

UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FARMACIA

Departamento de Microbiología II



TESIS DOCTORAL

Variabilidad natural en los sitios de procesamiento viral y en posiciones asociadas a resistencias antirretrovirales de la proteasa, retrotranscriptasa e integrasa en las variantes del VIH-1

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Teresa Llácer Delicado

Directores

África Holguín Ferbández,
Esther Torrecilla García

Madrid, 2018

Universidad Complutense de Madrid

Facultad de Farmacia

Departamento de Microbiología II



TESIS DOCTORAL

**“VARIABILIDAD NATURAL EN LOS SITIOS DE PROCESAMIENTO
VIRAL Y EN POSICIONES ASOCIADAS A RESISTENCIAS A
ANTIRRETROVIRALES DE LA PROTEASA, RETROTRANSCRIPTASA E
INTEGRASA EN LAS VARIANTES DEL VIH-1”**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR PRESENTADA POR

Teresa Llácer Delicado

DIRECTORES

Dra. África Holguín Fernández

Dra. Esther Torrecilla García

Madrid, 2017

La Dra. Esther Torrecilla García, Doctora por la Universidad Complutense de Madrid, adscrita al Servicio de Microbiología y Parasitología dentro del Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), y

La Dra. África Holguín Fernández, Doctora en Biología, investigadora del Sistema Nacional de Salud y responsable del Laboratorio de Epidemiología Molecular del VIH-1 del Servicio de Microbiología y Parasitología del Hospital Universitario Ramón y Cajal-IRYCIS

CERTIFICAN QUE:

TERESA LLÁCER DELICADO ha realizado bajo su dirección en el Laboratorio de Epidemiología Molecular del VIH-1 del Servicio de Microbiología del Hospital Universitario Ramón y Cajal-Instituto Ramón y Cajal de Investigaciones Sanitarias (IRYCIS), el trabajo que presenta para optar al grado de doctor, titulado:

“VARIABILIDAD NATURAL EN LOS SITIOS DE PROCESAMIENTO VIRAL Y EN POSICIONES ASOCIADAS A RESISTENCIAS A ANTIRRETROVIRALES DE LA PROTEASA, RETROTRANSCRIPTASA E INTEGRASA EN LAS VARIANTES DEL VIH-1”

Y para que así conste, firman la presente certificación en Madrid, 2017

Fdo.: Dra. África Holguín Fernández

Fdo.: Dra. Esther Torrecilla García



A mis padres

AGRADECIMIENTOS

«Dentro de veinte años estarás más decepcionado por las cosas que no hiciste que por aquellas que hiciste. Así que suelta amarras. Navega lejos del puerto seguro. Atrapa los vientos alisios en tus velas. Explora. Sueña. Descubre». Mark Twain (1835-1910)

Afortunadamente *in my life (Lennon and McCartney, 1965)* son muchas las personas a las que tengo que agradecer su apoyo y ayuda durante estos años tesiles. Como no soy de muchas palabras pero si de mucha música he dispuesto unos agradecimientos un poco cantarines y es que la música compone los ánimos descompuestos y alivia los trabajos que nacen del espíritu (*Cervantes, 1547*).

Mi enorme agradecimiento a todos, a mi familia, amigos y compañeros, que han puesto granitos, montones e incluso toneladas de arena para que llegara a tierra firme y pusiera fin a esta larga travesía que en muchos momentos parecía tratarse de un *long and winding road (McCartney et al., 1970)* y en otros tantos resultaba emocionante y alegre como un *beatiful day (Bono et al., 2000)*.

Me gustaría agradecer al Dr. Rafael Cantón el permitirme realizar esta tesis en el Servicio de Microbiología y Parasitología del Hospital Universitario Ramón y Cajal.

A mis directoras de tesis África Holguín y Esther Torrecilla, por su apoyo constante desde principio a fin, su enorme capacidad de trabajo y sacrificio. Sin ellas esta Tesis no habría sido posible. A la Dra. África Holguín, por darme la oportunidad de realizar la tesis, confiar en mí y transmitirme su interés, entusiasmo y voluntad inquebrantable en este proyecto. A la Dra. Esther Torrecilla por acompañarme durante todo el camino antes de ser directora de mi tesis y después, por su generosidad, su amistad y por afrontar nuestras jornadas maratónicas de trabajo siempre con una sonrisa hasta el crepúsculo (*Erentxun y Vasallo, 2001*).

A todos los compañeros que han pasado por el Laboratorio EpiMol y que lo integran en la actualidad, de todos he aprendido y pasado buenos ratos. A Gonzalo, Miguel, Leticia, Bárbara, Laura, Carolina, Sara, Lucia, Patri R., Marina. A Patri A por sus consejos prácticos y los music moments *like a Rolling Stone (Dylan, 1965)*.

A mis compañeras del Servicio de Micro Irene, Ana Sofi y María y también a Rocío, por su inestimable ayuda con los trámites tesiles y demás adyuvantes *under pressure (Queen and Bowie, 1981)* al final del proceso.

A mis EFIG (Isa, Cris, Lidia, David, Ana, Bego, Paloma) los mejores compañeros de residencia que he podido tener, espero que nuestra amistad perdure por lo menos otros diez años más *can't stop this thing we started (Adams, 1991)*.

A mis pediátrines de la AEMPS, Txus, Estela, Eduardo, Celia, gracias por aguantarme en el día a día, ayudarme en lo que habéis podido y apoyarme *walk on the wild side (Reed, 1972)*. A mis compañeras de ensayos clínicos y todos los compañeros de la AEMPS que de alguna u otra forma se han preocupado por mi estado tesil. A Dulci, Teresita, Patri, Lara, Luisa. A Ino por su generosa ayuda y empujón final, *dream on (Tyler, 1973)*.

A mis compañeros de la Berli, excelentes compañeros gracias por confiar en mí, vuestro cariño y amistad, casi nueve años después os seguís acordando de mí y yo de vosotros, *Miss you (Jagger and Richards, 1978)*. A Cris, Carolines, Esther Startis, Gemma, el Dr. LuisMi, Captain, Joquin, maester Mario gracias por vuestra amistad.

A mis amigos, todos ellos con un *heart of gold (Young and The Stray Gators, 1972)* y con ellos *a little help for my friends (Lennon and McCartney, 1967)* se queda corta, gracias por vuestra amistad y vuestro apoyo en esta travesía y en todas. A mis amigas valencianas, a Maripichi por estar siempre a mi lado incondicionalmente en todos los momentos, *don't let the sun go down on me (John, 1974; John and Michael 1991)*. A Adela por su positividad y entusiasmo, por nuestras charlas científicas *bohemian rhapsody (Mercury, 1975.)* A Inma y Lola *my generation (The Who, 1965)*. A mis amigas madrileñas de residencia. A Arantza, por su apoyo y generosidad en todos los momentos, su escucha activa y sus consejos sinceros. A Isa por toda su ayuda y soporte, su disponibilidad y su gran empujón final. A Celia por su ayuda y recomendaciones. A Vero por su ánimo y comprensión, a Rosa por estar pendiente, Ana Ries Baixes, Tere, Patri gracias por vuestros ánimos *what a difference a day makes (Cullum, 2003; Washington, 1959)*. A mis amigas madrileñas que aún en la distancia han seguido brindándome su apoyo y amistad. A Martuki, Belén, Pilar, Sara.

A mis padres, Rosa y Jesús, que me lo han dado todo pero sobre todo amor, *love theme (White ,1974)*. Su ejemplo de lucha, sacrificio, trabajo y superación ha sido un referente para mí. Papá *wish you were here (Gilmour and Waters, 1975)*. Mamá *la mujer que yo quiero (Serrat, 1971)*. A mis hermanos, Jesús y Josep. A mi Sofi, *hey Jude don't be afraid (McCartney and Lennon, 1968)*. A mi tía Llani, por su apoyo familiar en los momentos difíciles, por su ejemplo de humanidad, tolerancia y superación personal, *One (Bono et al., 1992)*.

Gracias a la música, los conciertos, a Bruce, las carreras, maratones, triatlones, desafíos Calleja, los viajes y escapadas, los bailes de bollywood y africanos, el body combat, las cervecitas de después del gym y los aperitivos de los domingos, *satisfy my soul (Carrack, 2000)*.

A todos y todo lo que me ha traído hasta aquí gracias

Con todo cariño, Tere

AGRADECIMIENTOS

ÍNDICE

ÍNDICE DE CONTENIDOS

	Página
ABREVIATURAS	1
RESUMEN	9
SUMMARY	15
INTRODUCCIÓN	21
EPIDEMIOLOGÍA DE LA INFECCIÓN POR VIH	23
CURSO NATURAL DE LA INFECCIÓN.....	24
ESTRUCTURA Y GENOMA DEL VIH.....	26
CICLO BIOLÓGICO DEL VIRUS	28
SITIOS DE PROCESAMIENTO PROTEICO PARA GENERAR LAS PROTEÍNAS GAG Y POL	30
UBICACIÓN Y PAPEL BIOLÓGICO DE LAS PROTEÍNAS GAG Y POL	33
DIVERSIDAD GENÉTICA DEL VIH	35
VARIANTES DEL VIH	36
GENERALIDADES DEL TRATAMIENTO ANTIRRETROVIRAL	40
EVALUACIÓN Y MONITORIZACIÓN DEL TRATAMIENTO ANTIRRETROVIRAL	41
RESISTENCIAS A ANTIRRETROVIRALES TRANSMITIDAS Y ADQUIRIDAS. IMPLICACIÓN CLÍNICA ..	44
VARIABILIDAD NATURAL EN RESIDUOS DE POL RELACIONADOS CON RESISTENCIA A ANTIRRETROVIRALES.....	48
GAG COMO DIANA TERAPEÚTICA	50
MUTACIONES EN GAG IMPLICADAS EN RESISTENCIAS A INHIBIDORES DE LA PROTEASA	50
MATERIAL Y METODOS	53
SECUENCIAS DEL VIH-1.....	55
DESCARGA DE SECUENCIAS DE VIH DE LA BASE DE DATOS GENBANK.....	55
DESCARGA DE SECUENCIAS CONSENSO DE CADA SUBTIPO, SUB-SUBTIPO Y RECOMBINANTE CRF DEL VIH-1	58
GENERACIÓN DE SECUENCIAS CONSENSO A PARTIR DEL ALINEAMIENTO DE SECUENCIAS DISPONIBLES EN GENBANK	59
ALINEAMIENTO Y EDITADO DE SECUENCIAS DE NUCLEÓTIDOS Y AMINOÁCIDOS	60
IDENTIFICACIÓN DE LOS RESIDUOS Y DOMINIOS CONSERVADOS ENTRE VARIANTES	61
IDENTIFICACIÓN DE POSICIONES EN GAG Y POL IMPLICADAS EN RESISTENCIA A FÁRMACOS ANTIRRETROVIRALES.....	61
PREDICCIÓN DE LA SUSCEPTIBILIDAD A ANTIRRETROVIRALES.....	62
ANÁLISIS ESTADÍSTICO.....	62
CAPÍTULO 1	63
VARIABILIDAD EN LOS SITIOS DE PROCESAMIENTO PROTEICO ENTRE GRUPOS, SUBTIPOS Y RECOMBINANTES DEL VIH-1.....	63

JUSTIFICACIÓN Y OBJETIVOS	65
MATERIAL Y MÉTODOS	66
RESULTADOS	68
DISCUSIÓN	78
 CAPÍTULO 2	 85
VARIABILIDAD NATURAL EN POSICIONES ASOCIADAS A RESISTENCIA A INHIBIDORES DE LA PROTEASA (MRD-IP) EN GAG Y PR DE DISTINTAS VARIANTES DEL VIH-1. EFECTO DEL TAR EN ALGUNAS MRD-IP EN GAG	85
JUSTIFICACIÓN Y OBJETIVOS	87
MATERIAL Y MÉTODOS	88
RESULTADOS	93
 CAPÍTULO 3	 111
VARIABILIDAD NATURAL EN POSICIONES DE LA RT ASOCIADAS A RESISTENCIA A INHIBIDORES DE LA TRANSCRIPTASA INVERSA EN LAS DISTINTAS VARIANTES DEL VIH-1 EN PACIENTES <i>NAIVE</i>	111
JUSTIFICACIÓN Y OBJETIVOS	113
MATERIAL Y MÉTODOS	115
RESULTADOS	127
 CAPÍTULO 4	 141
VARIABILIDAD NATURAL EN POSICIONES ASOCIADAS A RESISTENCIA A INHIBIDORES DE LA INTEGRASA (INI) EN PACIENTES <i>NAIVE</i> A INI	141
JUSTIFICACIÓN Y OBJETIVOS	143
MATERIAL Y MÉTODOS	144
RESULTADOS	149
DISCUSIÓN	152
 DISCUSIÓN	 155
 INTEGRADORA	 155
 CONCLUSIONES	 181
 BIBLIOGRAFÍA.....	187
 ANEXOS.....	213
 ANEXO I	215
 ESTRUCTURA DE LOS CRF DEL VIH-1	215
 ANEXO II	229

FÁRMACOS ANTIRRETROVIRALES AUTORIZADOS EN EUROPA AGRUPADOS POR FAMILIAS	229
ANEXO III	233
COMBINACIONES DE TAR DE INICIO RECOMENDADAS (GESIDA 2017)	233
ANEXO IV	237
PUBLICACIONES.....	237
ANEXO V	257
OTRAS PUBLICACIONES.....	257

ÍNDICE DE FIGURAS

	Página
Figura I.1. Curso natural de la infección por VIH.	25
Figura I.2. Estructura del virión (A) y genoma del VIH (B).	27
Figura I.3. Ciclo replicativo del VIH.	29
Figura I.4. Sitios de procesamiento proteico en precursores virales Gag y GagPol del VIH-1 (A) y secuencias de aa (B).	31
Figura I.5. Esquema de la clasificación de las variantes del VIH.	37
Figura I.6. Distribución mundial de los subtipos y algunas formas recombinantes del VIH-1.	39
Figura M.1. Descarga de secuencias de la Base de Datos del GenBank.	56
Figura M.2. Secuencia de la PR del VIH-1 en formato fasta.	57
Figura M.3. Descarga de las secuencias consenso del VIH-1 disponibles en GenBank.	59
Figura M.4. Alineamientos editados con MEGA en nt (A) o en aa (B).	60
Figura 1.1. Distribución de variantes VIH-1 del grupo M entre las 9.028 secuencias gag (A) y 3.906 pol (B).	69
Figura 1.2. Familias de recombinantes CRF del VIH-1 establecidas para este estudio.	70
Figura 1.3. Secuencias consenso en cada CS en variantes del VIH-1 con discrepancias respecto a la secuencia consenso de consensos de GenBank.	72
Figura 1.4. Conservación de los 11 sitios de procesamiento viral o CS, a nivel de aa, localizados en los precursores Gag y GagPol, en las variantes VIH-1 analizadas.	75
Figura 1.5. Conservación de los primeros y últimos sitios de procesamiento entre grupos del VIH-1.	76
Figura 2.1. Variabilidad natural en posiciones de Gag (A) y de PR (B) asociadas a MRD-IP en 2.141 secuencias de pacientes <i>naïve</i> en GenBank.	98
Figura 2.2. Origen geográfico de los 106 pacientes <i>naïve</i> y los 43 tratados con IP.	105
Figura 3.1. Clasificación de las MRD en la RT a ITIAN e ITINAN	119
Figura 3.2. Variabilidad natural en posiciones de RT asociadas a MRD-ITIAN (A) y de MRD-ITINAN (B) en 3.592 secuencias de pacientes <i>naïve</i> en las variantes VIH-1 con secuencias disponibles en GenBank.	128
Figura 4.1. Identificación de los criterios de búsqueda de las secuencias IN de estudio.	145

Figura 4.2. Variabilidad natural en cada posición MRD-INI entre las variantes VIH-1. 151

Figura D.1. Implicaciones diagnósticas y clínicas de las infecciones por variantes no-B del VIH-1. 158

Figura D.2. Polimorfismos naturales en Gag, RT, PR e IN asociados a MRD en gag y pol presentes de manera natural en $\geq 90\%$ de secuencias de las distintas variantes del VIH-1. 173

ÍNDICE DE TABLAS

	Página
Tabla I.1. Funciones de las proteínas Gag y Pol.	33
Tabla I.2. Longitud y ubicación de las proteínas Gag y Pol generadas durante la maduración viral.	34
Tabla I.3. Listado de mutaciones de resistencia a antirretrovirales en pacientes <i>naïve</i> (A) y tratados (B) por familias de fármacos.	46
Tabla I.4. Polimorfismos y resistencias naturales en variantes no-B del VIH-1.	49
Tabla 2.1. Grado de resistencia o score previsto por Stanford para cada MRD en la PR Primarias (A), Secundarias (B) y Combinaciones (C)	91
Tabla 2.2. MRD-IP en Gag considerados y polimorfismos naturales previamente descritos.	94
Tabla 2.3. Polimorfismos naturales en PR asociados a MRD-IP presentes de manera natural en las distintas variantes del VIH-1.	102
Tabla 2.4. Descripción de 149 pacientes infectados por VIH-1 con secuencia <i>gag</i> .	104
Tabla 2.5. Conservación de 4 CS en virus de distintas variantes en 149 pacientes infectados con y sin experiencia previa a IP.	106
Tabla 3.1. Grado de resistencia o score para ITIAN (A) o ITINAN (B) predicho por Stanford para cada MRD en la RT.	121
Tabla 3.2. Diferencias en MRD-ITIAN (A) e MRD-ITINAN (B) empleando el programa HIVdb de Stanford con las secuencias de GenBank.	133
Tabla 3.3. Diferencias en la identificación de la variante VIH-1 empleando el programa HIVdb de Stanford.	136
Tabla 4.1. Mutaciones de resistencia a INI y scores según Stanford.	147
Tabla 4.2. Mutaciones de resistencia a INI descritas en otras fuentes diferentes de Stanford.	148
Tabla D.1. Polimorfismos naturales en Gag, RT, PR e IN asociados a MRD en gag y pol presentes de manera natural en $\geq 90\%$ de secuencias de las distintas variantes del VIH-1.	164
Tabla D.2. MRD ausentes en las secuencias de pacientes <i>naïve</i> analizadas para esta Tesis	171

ÍNDICE DE ANEXOS

	Página
ANEXOS	
Anexo I. Estructura de los CRF del VIH-1.	217
Anexo II. Fármacos antirretrovirales autorizados en Europa agrupados por familias	231
Anexo III. Combinaciones de TAR de inicio recomendadas	235
Anexo IV. Publicaciones.	239
Anexo V. Otras Publicaciones.	261

ABREVIATURAS

aa	Aminoácido
ADN	Ácido desoxirribonucleico
AIDS	<i>Acquired immune deficiency syndrome</i> o síndrome de inmunodeficiencia adquirida
ANRS	Agence nationales de reserche sur le sida et les hépatites virales o Agencia Nacional Francesa para la Investigación del Sida y hepatitis virales
ARN	Ácido ribonucleico
ARNm	ARN mensajero
ARV	Antirretroviral(es)
CA	Proteína de la cápside o P24
CDC	<i>Centre for Disease Control and Prevention</i> o Centros para el Control y la Prevención de Enfermedades de E.E. U.U.
CDS	Secuencia codificante (Pol CDS)
CIBERESP	Centro de Investigación Biomédica en Red Epidemiología y Salud Pública
CRF	<i>Circulating recombinant forms</i> o Forma(s) recombinante(s) circulante(s)
CS	<i>Cleavage site(s)</i> o Sitio(s) de procesamiento proteico
CV	Carga Viral
ECDC	<i>European Centre for Disease Prevention and Control</i> o Centro Europeo para la Prevención y el Control de las Enfermedades
EMA	<i>European Medicines Agency</i> o Agencia Europea de Medicamentos
EpiMolVIH	Laboratorio de Epidemiología Molecular del VIH
EE. UU.	Estados Unidos de América
FDA	<i>US Food and Drug Administration</i> o Agencia de medicamentos alimentos estadounidense
FV	Fracaso virológico
GCS	Sitios de procesamiento proteico del precursor Gag
GeSIDA	Grupo de estudio del Sida
HSH	Hombres que tienen sexo con hombres
IAS-USA	<i>International AIDS Society</i> o Sociedad Internacional sobre el Sida de EE. UU
IN	Integrasa
IP	Inhibidores de la proteasa
INI	Inhibidores de la integrasa

INI-R	Resistencia a inhibidores de la integrasa
INSTI	Inhibidores de la transferencia de cadenas de la integrasa
IRYCIS	Instituto Ramón y Cajal de Investigación Sanitaria
ITIAN	Inhibidores de la transcriptasa inversa análogos de nucleósido o nucleótido
ITINAN	Inhibidores de la transcriptasa inversa no análogos de nucleósido
LANL	<i>Los Alamos National Laboratory</i> o Laboratorio Nacional de Los Álamos, EE. UU.
log	Logaritmo decimal
LTNP	<i>Long term non progressors</i> o progresores lentos (a Sida)
LTR	Repeticiones largas terminales
MA	Proteína de la matriz o P17
MRD	Mutaciones de resistencia
MRD-INI	Mutaciones de resistencia a inhibidores de la integrasa
MRD-IP	Mutaciones de resistencia a inhibidores de la proteasa
MRD-ITIAN	Mutaciones de resistencia a inhibidores de la transcriptasa inversa análogos de nucleósido o nucleótido
MRD-ITINAN	Mutaciones de resistencia a inhibidores de la transcriptasa inversa no análogos de nucleosido
MEGA	<i>Molecular Evolutionary Genetics Analysis</i>
MRT	Mutaciones de resistencia transmitidas
MRA	Mutaciones de resistencia adquiridas
Nº	Número
NC	Proteína de la nucleocápsida o P7
NCBI	<i>National Center for Biotechnology Information</i> o Centro Nacional de Información Biotecnológica, EE. UU.
NIAID	<i>National Institute of Allergy and Infectious Diseases</i> o Instituto Nacional de Alergia y Enfermedades Infecciosas, EE. UU.
NIH	<i>National Institutes of Health</i> o Institutos Nacionales de Salud, EE. UU.
nt	Nucleótido
OMS	Organización Mundial de la Salud
ONUSIDA	Programa Conjunto de las Naciones Unidas sobre el VIH/Sida
PCR	Reacción en cadena de la polimerasa
PID	Personas que se inyectan drogas

PNS	Plan Nacional sobre el Sida
PR	Proteasa
RIS	Red Española de Investigación en Sida
RT	Retrotranscriptasa
RyC	Ramón y Cajal
SEFH	Sociedad Española de Farmacia Hospitalaria
Sida	Síndrome de la inmunodeficiencia adquirida
TAM	Mutaciones de análogos de timidina
SINVIH	Sistema de información sobre nuevos diagnósticos del VIH
TAR	Tratamiento antirretroviral
TI	Transcriptasa inversa
TFP	Proteína marco de lectura (<i>transframe protein</i>)
UDVP	Usuarios de drogas por vía parenteral
UE	Unión Europea
URF	<i>Unique recombinant forms</i> ; Forma(s) recombinante(s) única(s)
VIH	Virus de la inmunodeficiencia humana
VIS	Virus de la inmunodeficiencia del simio
WHO	<i>World Health Organization</i>

FÁRMACOS ANTIRRETROVIRALES FRENTE AL VIH

Inhibidores de la transcriptasa inversa análogos de nucleós(t)ido (ITIAN)

3TC	Lamivudina
ABC	Abacavir
AZT	Zidovudina
d4T	Estavudina
FTC	Emtricitabina
TDF	Tenofovir disoproxil fumarato
TAF	Tenofovir alafenamida

Inhibidores de la transcriptasa inversa no análogos de nucleósido (ITINAN)

EFV	Efavirenz
ETV	Etravirina
NVP	Nevirapina
RPV	Rilpivirina

Inhibidores de la proteasa (IP)

ATV	Atazanavir
DRV	Darunavir
FPV	Fosamprenavir
IDV	Indinavir
LPV	Lopinavir
RTV	Ritonavir
SQV	Saquinavir
TPV	Tipranavir

Potenciadores farmacocinéticos (de atazanavir, lopinavir, darunavir o elvitegravir; se emplean en coadministración con los IP citados)

RTV (/r)	Ritonavir
COBI (/r)	Cobicistat

Inhibidores de la integrasa (INI o INSTI)

RAL	Raltegravir
EVG	Elvitegravir
DTG	Dolutegravir

Inhibidores de la fusión (IFI)

ENF	Enfuvirtida
------------	-------------

Inhibidores CCR5

MVC	Maraviroc
------------	-----------

AMINOÁCIDOS

A	Alanina
C	Cisteína
D	Ácido aspártico
E	Ácido glutámico
F	Fenilalanina
G	Glicina
H	Histidina
I	Isoleucina
K	Lisina
L	Leucina
M	Metionina
N	Asparagina
P	Prolina
Q	Glutamina
R	Arginina
S	Serina
T	Treonina
V	Valina
W	Triptófano
Y	Tirosina

RESUMEN

“Hay que amar la vida, y no solo la nuestra”

Emilio Lledó (1927)

El virus de la inmunodeficiencia humana tipo 1 (VIH-1) presenta alta variabilidad genética, debida a su alta tasa de replicación, de mutación y de recombinación, describiéndose numerosas variantes. Basándose en la homología genética de sus secuencias, el VIH-1 se clasifica en cuatro grupos: M (principal), O (*outlier*), N (non-M, non-O) y P. A su vez, el grupo M se subdivide en 9 subtipos (A, B, C, D, F, G, H, J, K), en al menos 88 recombinantes circulantes (CRF), y múltiples recombinantes únicos. Los recombinantes se están expandiendo en la pandemia (incluido en Europa), principalmente en lugares donde circulan simultáneamente numerosas variantes diferentes. El subtipo B es la variante predominante en Europa y Estados Unidos, pero los subtipos del VIH-1 distintos al subtipo B y recombinantes (variantes no-B) causan cerca del 90% de los 37 millones de infecciones por VIH.

La extraordinaria variabilidad genética del VIH tiene implicaciones clínicas y epidemiológicas importantes. La rápida aparición y selección de mutaciones de resistencia (MRD) a fármacos antirretrovirales (ARV) implica un desafío para la terapia antirretroviral (TAR). Además, las distintas variantes del VIH presentan en su genoma de manera natural (en ausencia de ARV) sustituciones de aminoácidos (aa) fijadas durante la evolución, incluyendo en residuos de proteínas virales donde aparecen los MRD, como la proteasa (PR), retrotranscriptasa (RT) e integrasa (IN). Estos polimorfismos naturales pueden acelerar la emergencia de MRD, inducir vías alternativas de resistencia, reducir la barrera genética de algunos fármacos, afectar a la afinidad de unión fármaco-sustrato y disminuir la susceptibilidad a ARV, entre otros. Además, se sabe que polimorfismos en los sitios de procesamiento proteico (*cleavage sites* o CS) influyen, tanto el procesamiento proteico del VIH-1 por la PR viral requerido para la maduración e infectividad viral, como en la respuesta a ARV inhibidores de la PR (IP). Sin embargo, existen pocas publicaciones sobre variabilidad natural en CS y MRD asociada a las variantes del VIH-1, sobre todo en recombinantes.

Por ello, el objetivo principal de esta Tesis fue evaluar el grado de conservación, identificar los polimorfismos naturales presentes en más del 90% de secuencias virales de cada variante analizada en ausencia de tratamiento antirretroviral (TAR) y los aa totalmente conservados en regiones virales con importancia biológica y terapéutica. Estas incluían 11 CS en los precursores virales Gag y GagPol y los residuos de proteínas virales

Gag (dentro y fuera de los CS), PR, RT e IN donde aparecen las MRD a las principales familias de ARV: IP, inhibidores de la RT análogos de nucleósido o nucleótido (ITIAN), y no análogos de nucleósido (ITINAN) e inhibidores de la IN (INI). Para ello se emplearon varios miles de secuencias virales recuperadas del GenBank de pacientes *naïve* a ARV de numerosas variantes del VIH-1, incluyendo grupos, subtipos y el mayor número de CRF analizados hasta la fecha.

En el primer capítulo de esta Tesis empleamos 9.028 secuencias *gag* y 3.906 secuencias *pol* del GenBank de pacientes *naïve*, cuantificando el grado de conservación viral en 11 CS. Éste difirió entre variantes del VIH-1 y CS, y fue significativamente mayor en el grupo M que en los grupos no-M. Determinamos los CS que presentaban mayor conservación en subtipos que en recombinantes, y viceversa. Identificamos los polimorfismos naturales asociados a cada variante y los 52 aa altamente conservados en todas las secuencias. Los sitios de procesamiento tardío en los precursores Gag y GagPol virales presentaron una conservación significativamente mayor que el primer CS en los 4 grupos del VIH-1, sugiriendo que el nivel de conservación encontrado en los CS es distinto según las etapas de procesamiento. Definimos las secuencias consenso o mayoritarias en cada uno de los 11 CS para cada grupo, subtipo y CRF analizado. También observamos cómo la presencia de IP parecía no interferir significativamente en el grado de conservación de la mayoría de los 4 CS analizados empleando virus de variantes diferentes recuperados de 149 muestras clínicas en ausencia y presencia de IP. Estos resultados podrían ayudar al diseño de inhibidores de maduración nuevos y de vacunas dirigidas a los CS.

En el capítulo dos analizamos la variabilidad natural en posiciones de Gag y PR asociadas a resistencia a IP (MRD-IP) en 2.141 secuencias GagPol, en el capítulo tres las asociadas a resistencia a ITIAN (MRD-ITIAN) e ITINAN (MRD-ITINAN) empleando 3.592 secuencias RT y en el cuarto capítulo aquellas relacionadas con resistencia a INI (MRD-INI) en 6.706 secuencias IN. En todos los casos identificamos los residuos asociados a MRD que podrían considerarse polimorfismos naturales asociadas a grupo, subtipo o CRF incluidos al estar presentes en el 90% o más de las secuencias analizadas de cada variante de estudio en ausencia de TAR. También identificamos todas las MRD en Gag, PR, RT e IN ausentes en todos los virus analizados de pacientes *naïve* y que, por tanto, sólo se

seleccionarían en presencia de TAR. Observamos que la exposición a IP no afectaba al grado de conservación de 11 MRD-IP ubicadas en 4 CS tras analizar secuencias virales de muestras clínicas de 149 pacientes infectados por distintas variantes del VIH-1.

Encontramos que la presencia de MRD primarias a IP, ITIAN, ITINAN e INI fue rara en pacientes *naïve*, estando presentes sólo en un bajo porcentaje de secuencias de algunas variantes, excepto la MRD-ITINAN Y181C que aparecía en todas las variantes del grupo O analizadas. Sin embargo, demostramos que un gran número de aa considerados previamente como MRD secundarias, accesorias o raras a las distintas familias de ARV son, en realidad, polimorfismos naturales de grupos, subtipos o CRF del VIH-1 concretos, presentes incluso en ausencia de tratamiento. El número de polimorfismos naturales asociados a MRD difirió entre variantes, y esas combinaciones de MRD podrían considerarse marcadores genéticos de las variantes que los presentan. La presencia natural (en ausencia de terapia) de polimorfismos naturales específicos de variante previamente asociados a resistencia a las principales familias de fármacos ARV, refuerza la importancia de tener en cuenta la variante viral en el seguimiento clínico del paciente VIH en TAR. Este trabajo también tiene potencial aplicabilidad en el desarrollo de nuevos fármacos ARV y en la identificación de epitopos conservados para el desarrollo de vacunas

SUMMARY

“La verdadera vida tiene lugar en nuestro interior”

Paul Auster (1947)

Human immunodeficiency virus type 1 (HIV-1) shows a high genetic diversity due to its high replication, mutation and recombination rates. A large number of HIV-1 variants have been described. Based on viral sequences homology HIV-1 has been classified into four groups: M (main), O (*outlier*), N (non-M, non-O) and P. HIV-1 Group M is subdivided into 9 subtypes (A–D, F–H, J, K), at least 88 circulating recombinant forms (CRF) and multiple unique recombinants (URF). Recombinants are expanding in the pandemic (including in Europe), mainly in setting with cocirculation of different HIV-1 variants. HIV-1 subtype B predominates in Europe and United States, but subtypes different than subtype B and recombinants (HIV-1 non–B variants) cause near 90% of HIV infections worldwide.

The high genetic HIV-1 variability has epidemiological and clinical consequences, and the fast appearance and selection of antiretroviral drug resistance mutations (MRD) represents a challenge for antiretroviral therapy (ART). Different HIV variants fix during viral evolution amino acid (aa) changes along their genome in the absence of antiretroviral (ARV), including in viral protein residues previously associated to ARV resistance in the protease (PR), retrotranscriptase (RT) and integrase (IN). Those HIV-variant specific-natural polymorphisms can accelerate the MRD selection, drive the selection of alternative resistance pathways, reduce the genetic barrier to certain drugs, affect the ARV-virus affinity and reduce the ARV susceptibility, among others. Moreover, it is known that polymorphisms at cleavage sites (CS) can influence Gag and Pol proteins processing by the viral protease (PR), restore viral *fitness* and influence the virological outcome of specific antiretroviral drugs of PR inhibitoris (PI). However, data of HIV-1 variant associated CS natural variability is scarce.

Thus, the main objective of the Thesis has been to evaluate the conservation rate, to identify the natural polymorphisms present in 90% or more of viral sequences in each analyzed variant in the absence of ART and the most/completely conserved aa residues in HIV-1 genomic regions with biological and clinical importance. They include the 11 CS and the amino acid (aa) residues of viral proteins Gag, PR, RT and IN where are located the MRD to the most used ARV families: PR inhibitors (PI), RT inhibitors analogs of nucleosides and nucleotides (NRTI), and non-analogous of nucleosides (NNRTI) and IN inhibitors (INI). For that purpose we used several thousand of viral sequences retrieved

from GenBank database from ARV *naïve* patients, ascribed to a large panel of HIV-1 variants, including groups, subtypes, and the highest number of different circulant recombinants or CRF reported to date

In the first chapter of the Thesis we used 9.028 *gag* y 3.906 *pol* sequences from GenBank ascribed to a large number of different HIV-1 variants from *naïve* patients. We quantifying the conservation rate in 11 CS across HIV-1 variants, being significantly higher in group M variants vs. non-M groups. We determined the most and the least conserved CS across variants, defining the specific CS showing more conservation in subtypes than recombinants and *viceversa*. We identified the variant-specific natural polymorphisms in each CS and the 52 amino acids (aa) highly conserved in the 11 CS at Gag and GagPol precursors. The late processing sites at Gag and GagPol precursors presented a significantly higher conservation vs. the first CS in the 4 HIV-1 groups. We defined the aa consensus sequence in each HIV-1 variant for each 11 CS. We also observed that the PI exposure could not affect the conservation rate of some of CS in viral sequences from viral specimens collected from clinical specimens from patients carrying different HIV-1 variants. These results could help for the future design of both novel antiretroviral agents acting as maturation inhibitors as well as for vaccine targeting CS.

In chapter 2, we analyzed the natural variability at Gag and PR residues associated to PI resistance (MRD-PI) in 2.141 GenBank GagPol sequences, in chapter 3 those associated to NRTI (MRD-NRTI) and NNRTI (MRD-NNRTI) resistance using 3.592 GenBank RT sequences and in chapter 3 those related to INI resistance (MRD-INI) in 6.706 GenBank IN sequences representing multiple HIV-1 variants. We identified all specific aa residues associated to MRD that could be considered as natural specific-variant polymorphisms for each HIV-1 group, subtype and analyzed CRF, present in 90% or more analyzed sequences of the corresponding variant in the absence of ART. We also located all MRD in Gag, PR, RT and IN absent in all analyzed GenBank sequences from naïve patients and, accordingly, only would be selected after ARV exposure. We also observed that IP exposure did not affect the aa conservation rate in 11 MRD-IP located in 4 CS within after the analysis of viruses from clinical specimens and derived from 149 patients carrying different HIV-1 variants.

We also found that the presence of primary or major MRD to PI, NRTI, NNRTI and INI was rare in *naïve* patients, only being present in a low rate of sequences from certain HIV-1 variants, except the MRD-NNRTI Y181C which was in all analyzed RT sequences from HIV-1 group O. However, we demonstrated that a high number of some MRD considered secondary, accessory or rare to the different ARV families were, in fact, natural polymorphisms of specific HIV-1 groups, subtypes and CRF, being present even in the absence of ARV. The number of natural polymorphisms at MRD varied across variants, and those polymorphisms could be considered as genetic markers of these specific variants. The natural presence (in absence of ARV) of polymorphisms previously associated to ARV resistance reinforces the importance of considering the HIV-1 variant during the clinical follow up of HIV-1 infected patients on antiretroviral therapy. This work also has potential applicability in the development of new ARV drugs and identification of conserved epitopes for the vaccines development.

INTRODUCCIÓN

“El secreto de salir adelante es empezar”

Mark Twain (1835-1910)

EPIDEMIOLOGÍA DE LA INFECCIÓN POR VIH

El virus de la inmunodeficiencia humana (VIH) es el agente infeccioso causante del síndrome de inmunodeficiencia adquirida (Sida). El VIH continúa siendo un problema de salud pública mundial importante. Según el último informe del Programa Conjunto de las Naciones Unidas sobre el VIH/Sida (ONUSIDA), publicado en Junio de 2016, se calcula que a nivel global en 2015 había 36,7 millones de personas infectadas por VIH, 2,1 millones de nuevas infecciones y 1,1 millones de muertes por causas relacionadas con el VIH (Global AIDS update 2016). África Sub-Sahariana es la región más afectada a nivel mundial, con 25,6 millones de personas viviendo con VIH en 2015 y donde ocurren un 66% del total de las nuevas infecciones a nivel mundial. Aunque el número (nº) de nuevas infecciones por VIH ha disminuido en los últimos años (HIV/AIDS. WHO 2016), todavía existe una tasa elevada de nº de nuevas infecciones y muertes relacionadas con Sida cada año (*AIDS by the numbers* 2016). Además, 22 de los 36,7 millones de personas infectadas no tienen acceso al tratamiento antirretroviral (TAR) para controlar la replicación viral (*On the fast track to end AIDS* 2016).

En Europa el nº de casos de Sida sigue reduciéndose, gracias al uso y eficacia del TAR, excepto en algunos países de Europa del Este donde continúa incrementándose. En 2015, se comunicaron 29.747 casos diagnosticados de infección por VIH en 31 países de la Unión Europea. La proporción más alta, 42,2%, se notificó entre los hombres que mantienen relaciones sexuales con hombres (HSH), seguida por la del contacto heterosexual, 32,0% y un 4,2% en personas que se inyectan drogas (PID) (ECDC 2015).

En España, desde el inicio de la epidemia hasta el año 2015 se han notificado un total de 85.720 casos de Sida (SINIVIH 2016). En 2015, se notificaron 3.428 nuevos diagnósticos de VIH y 497 casos de Sida. La vía sexual es el modo de transmisión principal en los nuevos diagnósticos de VIH y la transmisión entre HSH es la mayoritaria (53,6%), seguida de la heterosexual (25,4%) y en personas que se inyectan drogas (PID), 2,8%. El 85,9% eran hombres y la mediana de edad fue de 36 años. El 30,3% de los nuevos diagnósticos de infección por el VIH se realizó en personas originarias de otros países. El 46,5% global de los nuevos diagnósticos presentaron diagnóstico tardío. Las tasas de nuevos diagnósticos de VIH son similares a las de otros países de la región Europea de la

OMS, aunque superiores a la media de los países de Europa Occidental y de la Unión Europea (SINIVIH 2016).

CURSO NATURAL DE LA INFECCIÓN

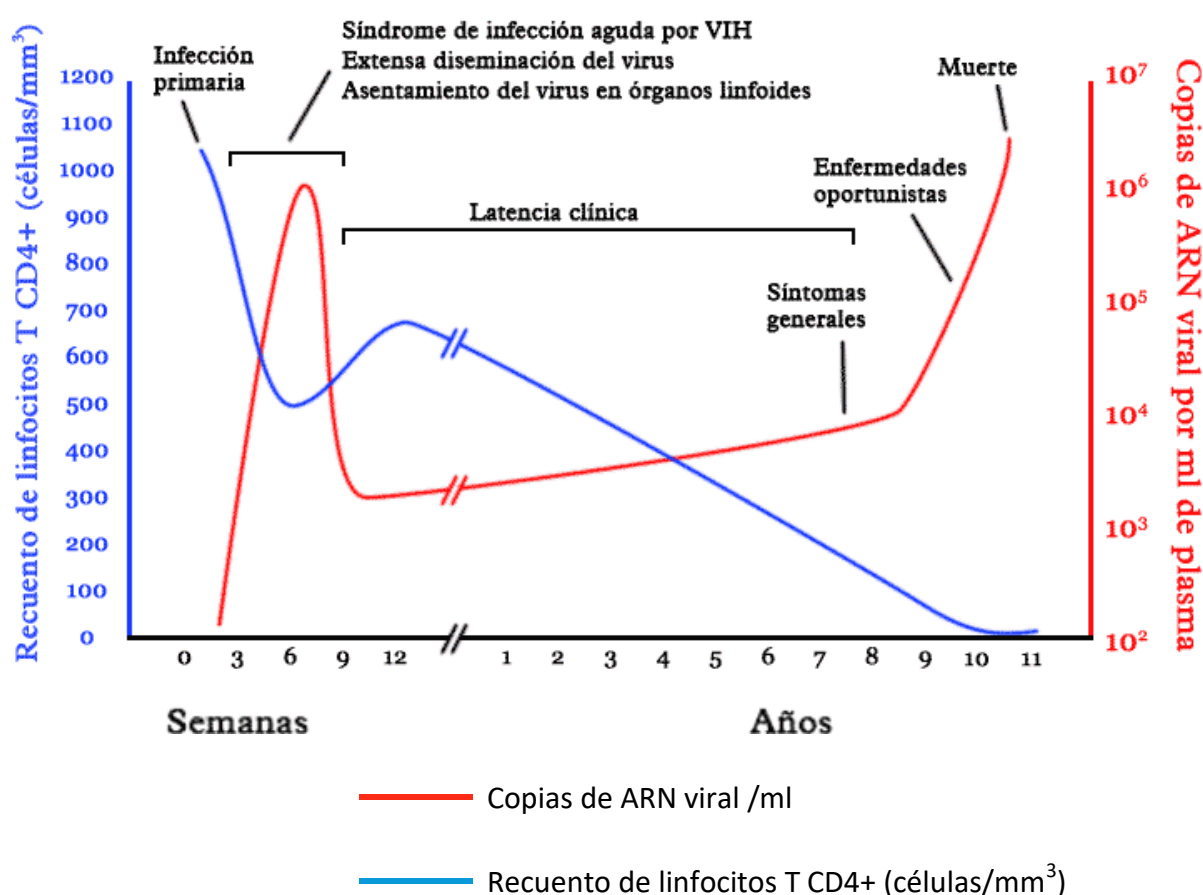
La infección por VIH transcurre como una infección viral persistente que causa replicación viral y enfermedad crónica. El curso natural de la infección por VIH-1 viene determinado por una serie de interacciones complejas entre el huésped y el virus (Liu *et al.*, 1997). Se pueden diferenciar fundamentalmente tres fases de la infección por el VIH, identificadas por un conjunto de síntomas e indicadores clínicos (**Figura I.1**).

La fase aguda, primoinfección o infección primaria transcurre entre el momento del contagio y la aparición de anticuerpos frente al VIH (4 semanas). Durante un “periodo ventana” (1-3 semanas) se produce una replicación masiva del virus dentro del organismo hasta alcanzar niveles en plasma o carga viral (CV) muy elevadas (varios millones de copias/ml plasma), infectando tanto a las células expuestas como a los ganglios linfáticos. El nº de linfocitos CD4 circulantes disminuye. Clínicamente, puede ser asintomática o con síntomas menores inespecíficos que no se identifican como relacionados con VIH, al ser similares a un síndrome gripal (con adenopatías y *rash* cutáneo) o a una mononucleosis infecciosa aguda. A continuación, a las pocas semanas o meses, la infección se encuentra en la fase crónica o de “latencia clínica”, coincidiendo con la aparición de la respuesta inmune y producción de anticuerpos (seroconversión). En esta fase el portador es asintomático o presenta complicaciones menores que, a menudo, no se identifican como relacionadas con el VIH-1, ya que el sistema inmune tiene una gran capacidad para regenerar las células destruidas por el virus. Sin embargo, la infección viral termina por desgastar al sistema inmunológico y aparece la fase final de la infección por VIH o etapa Sida, donde aumentan las tasas de replicación del virus, disminuyendo la capacidad de reacción del sistema inmune. Así, el sujeto es presa potencial de numerosas infecciones oportunistas y neoplasias que le pueden conducir a la muerte, si la infección no se controla por TAR.

La progresión de la enfermedad puede variar debido a factores asociados al virus y a factores genéticos e inmunológicos del hospedador (Merindol y Berhous, 2015). En ausencia de TAR, la mayoría (80-85%) de los pacientes desarrollan Sida en 5-10 años por

una caída progresiva de sus linfocitos T CD4+, siendo conocidos como progresores típicos. Sin embargo, según la velocidad de progresión de la infección, existen progresores rápidos, que progresan a Sida en pocos años; y progresores lentos o LTNP (*long term non progressors*), que permanecen clínicamente asintomáticos y/o inmunológicamente estables, con un recuento de linfocitos T CD4+ normal durante al menos 8 años sin TAR (Rodríguez-Da Silva *et al.*, 2015; GeSIDA 2017). Dentro de los LTNP, existe un grupo de pacientes, denominados controladores de élite, que controlan la viremia a niveles indetectables sin TAR y que representan en torno al 1% de infectados (Crowell *et al.*, 2015).

Figura I.1. Curso natural de la infección por VIH



ESTRUCTURA Y GENOMA DEL VIH

Es un virus ARN clasificado dentro de la familia de los retrovirus humanos (*Retroviridae*) perteneciente al género *Lentivirus* (Barré-Sinoussi *et al.*, 1983). Su origen se remonta a múltiples transmisiones zoonóticas del virus de la inmunodeficiencia del simio (VIS) desde diversos primates no humanos que tuvieron lugar en África central y occidental a principios del siglo XX (Worobey *et al.*, 2008).

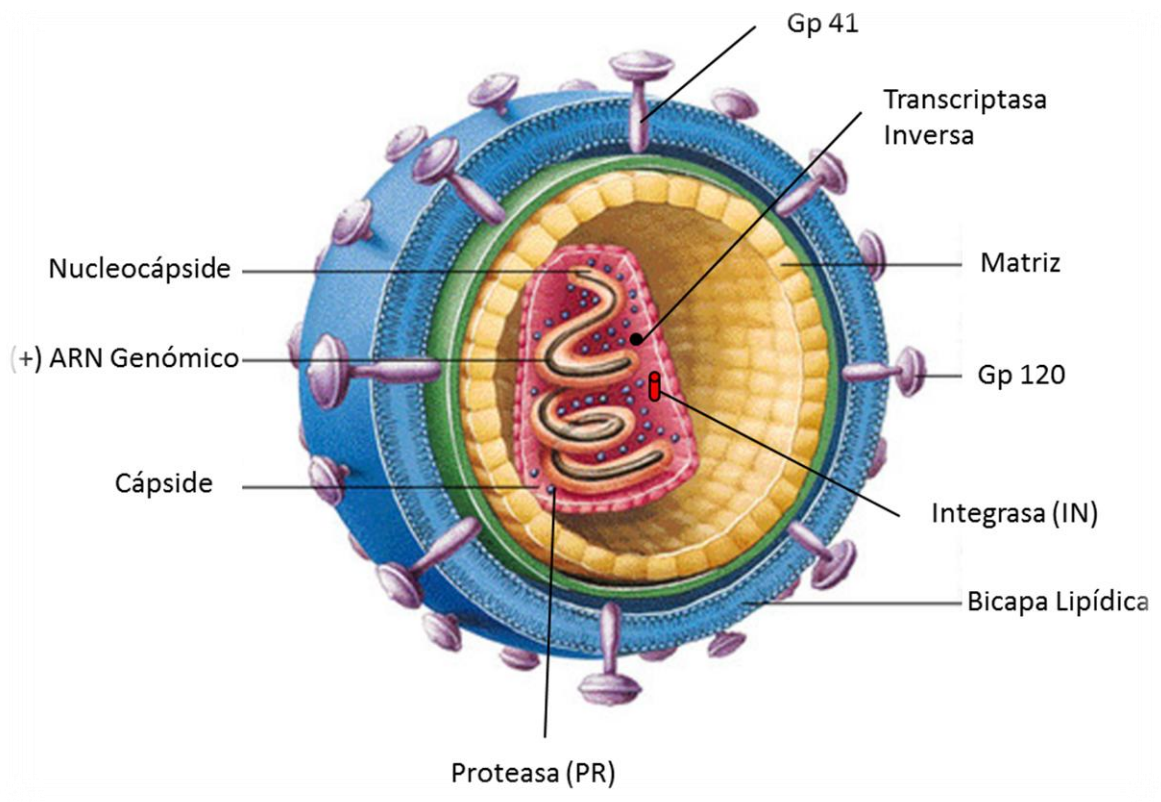
El VIH tiene forma esférica, con un diámetro de 100-120 nanómetros. El virión está constituido por una envoltura externa o bicapa lipídica, tomada de la membrana de la célula humana infectada durante el proceso de salida de nuevas partículas virales por gemación. En esta envoltura se encuentra la glicoproteína viral gp120, unida a la glicoproteína transmembrana gp41 en forma de trímeros y algunas proteínas de la célula huésped, como antígenos de histocompatibilidad de clase I y II. Las glicoproteínas gp120 son clave en la interacción con receptores celulares (tropismo) y en la evasión inmune. Bajo la envoltura se sitúa la matriz proteica (MA), y en el interior se encuentra la cápside icosaédrica (CA). Dentro de ésta se localizan, tanto las enzimas virales, como el material genético del virus constituido por dos cadenas sencillas de ARN de polaridad positiva de aproximadamente 9,8kb asociadas a las proteínas de la nucleocápside (NC) (Muesing *et al.*, 1985).

El genoma del VIH contiene tres genes principales comunes a todos los retrovirus: *gag*, *pol* y *env*. Adicionalmente, el VIH tiene genes encargados de codificar los componentes de la partícula vírica y de regular la expresión de los mismos. El gen *gag* codifica principalmente las proteínas estructurales que conforman la MA, la CA y la NC. El gen *pol*, codifica las proteínas virales: proteasa (PR), transcriptasa inversa (TI) o retrotranscriptasa (RT) e integrasa (IN) que participan en la maduración viral, síntesis de ADN a partir del ARN del virus y en su integración en el genoma celular, respectivamente. Estas proteínas son esenciales para el ciclo viral y actúan como dianas de fármacos antirretrovirales (ARV). El gen *env* codifica el precursor gp160 que se procesa a las glicoproteínas de la envuelta gp120 y gp41 por la acción de proteasas celulares. Además, en el genoma del VIH-1 existen otros seis genes adicionales (*tat*, *rev*, *nef*, *vif*, *vpr* y *vpu*) que codifican proteínas que regulan la expresión de genes virales o representan factores

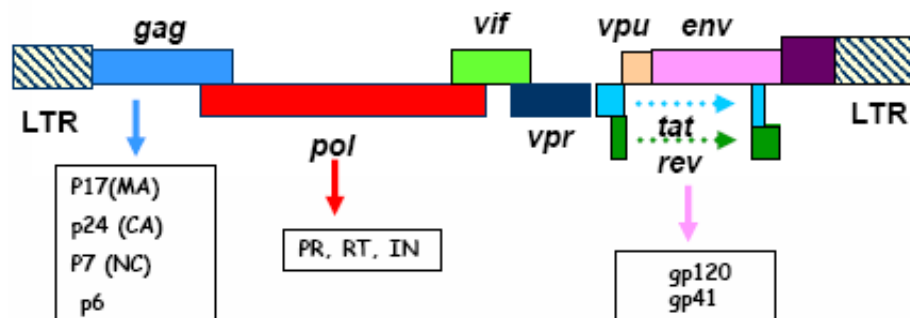
de virulencia. En ambos extremos del genoma, se encuentran secuencias largas repetidas (LTR), no codificantes, que permiten la circularización e integración del genoma viral en el genoma celular (Gallo *et al.*, 1988; Emerman y Malim, 1998). La estructura y organización genética del VIH-1 se muestran en la **Figura I.2.**

Figura I.2. Estructura del virión (A) y genoma del VIH (B)

(A)



(B)



CICLO BIOLÓGICO DEL VIRUS

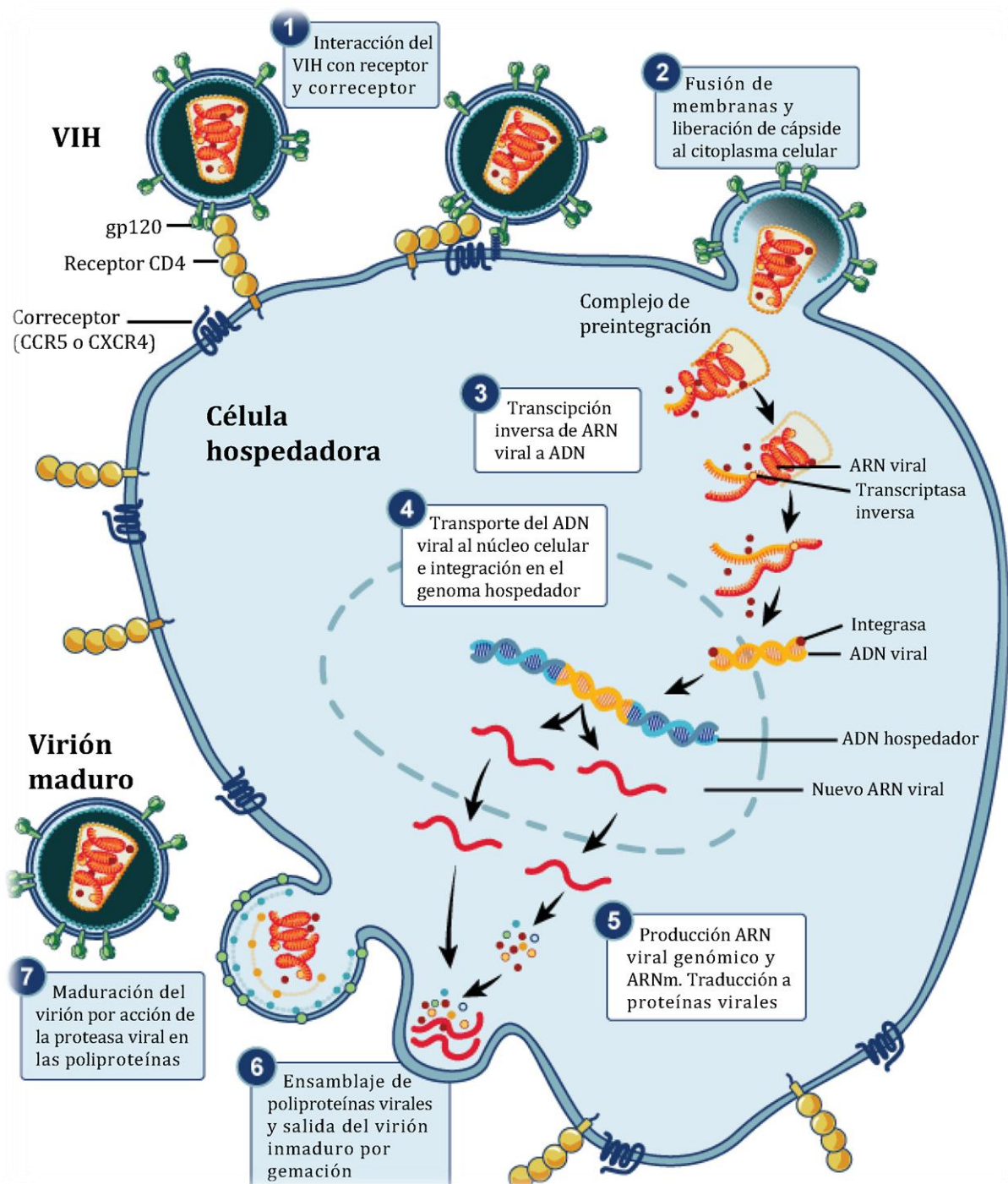
En el ciclo replicativo del VIH se resumen las siguientes etapas (**Figura I.3**):

- 1) Interacción del virus con la célula diana (linfocitos T CD4+ principalmente) por la unión de la glicoproteína gp120 de la envoltura viral con el receptor celular CD4 y un receptor de quimiocinas (CCR5 o CXCR4) que actúa como correceptor.
- 2) Fusión de la envuelta del virión con la membrana celular y liberación al citoplasma celular de la cápside viral (CA). La CA se desensambla y libera el genoma viral en el citoplasma.
- 3) Transcripción inversa del ARN genómico viral y formación del ADN complementario de doble cadena mediado por la RT. Este ADN proviral sintetizado se acopla a una serie de factores celulares y virales formando el complejo de preintegración.
- 4) Transporte de este ADN complejo al núcleo celular donde se integra en el genoma de la célula hospedadora por la acción de la IN viral y las secuencias LTR, constituyendo la forma proviral del VIH.

Una vez integrado, el VIH puede permanecer latente, replicarse de forma controlada o experimentar una replicación masiva, con el consiguiente efecto citopático sobre la célula infectada.

- 5) Transcripción de los genes virales integrados y procesamiento de los transcritos para dar lugar a ARN genómico y ARN mensajero (ARNm) virales. Traducción de los ARNm a poliproteínas precursoras virales en el citoplasma.
- 6) Ensamblaje de las proteínas virales y salida del virión por gemación, arrastrando parte de la membrana de la célula huésped.
- 7) Maduración por acción de la PR viral que procesa las poliproteínas precursoras para formar el virión infeccioso (Levy, 1993).

Figura I.3. Ciclo replicativo del VIH



Adaptado de *National Institute of Allergy and Infectious Diseases (NIAID)*, EE.UU. HIV Replication Cycle (<http://phil.cdc.gov/phil/details.asp ID#: 18162>).

SITIOS DE PROCESAMIENTO PROTEICO PARA GENERAR LAS PROTEÍNAS GAG Y POL

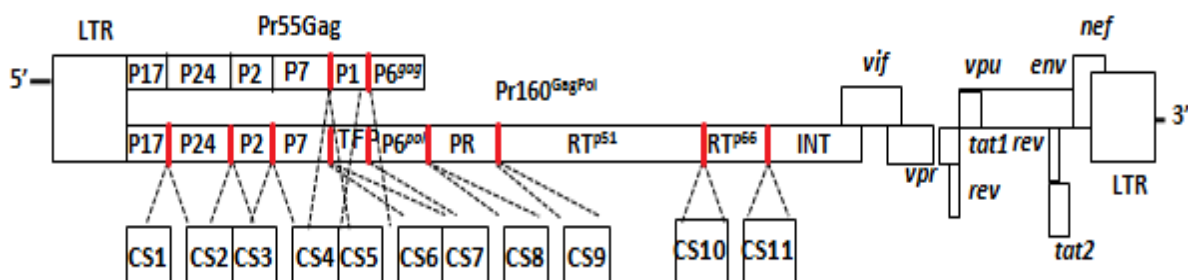
Las principales proteínas estructurales Gag del VIH-1 (P17 o MA, P24 o CA, P7 o NC y dos espaciadores llamados SP1 y SP2) se sintetizan a partir de una serie de reacciones proteolíticas. Éstas están mediadas por la PR viral y se realizan en unos sitios de corte específicos llamados sitios de procesamiento proteico ubicados en las poliproteínas precursoras Gag (Pr55^{Gag}) y GagPol (Pr160^{GagPol}). Ésta última poliproteína se genera gracias a un marco de lectura ribosomal adicional durante la traducción del precursor Gag (Waheed y Freed, 2012). Una sexta proteína producida a partir del precursor Gag, P6^{Gag}, juega un papel esencial en la liberación del virus de las membranas de la célula infectada (de Oliveira *et al.*, 2003). La P6^{Gag} es responsable de la incorporación de la proteína accesoria *vpr* al virión y de la interacción con la membrana de la célula que hace posible la gemación y la salida de la célula infectada (Göttlinger *et al.*, 1989). El procesamiento o maduración del precursor GagPol genera: las proteínas Gag estructurales (P17, P24, P7), la proteína *transframe* (TFP), P6^{Pol}, y las tres enzimas virales Pol (PR, RT e IN) (de Oliveira *et al.*, 2003; Torrecilla *et al.*, 2014) (**Figura 1.2.B**). Las proteínas Pol son esenciales para el ciclo viral y representan las principales dianas del TAR: inhibidores de la PR (IP), inhibidores de la TI análogos de nucleós(t)ido (ITIAN), inhibidores de la TI no análogos de nucleósido (ITINAN) e inhibidores de la IN (INI) también llamados inhibidores de la transferencia de cadenas de la IN (INSTI).

Estas reacciones de procesamiento de las poliproteínas Gag y GagPol del VIH-1 por la PR viral ocurren tardíamente en el ciclo de vida viral, una vez que el virión ha salido por gemación de la célula infectada, ya que la PR viral se activa durante la salida del virus. Este proceso proteolítico es altamente específico, está regulado de manera temporal y es esencial para la producción de partículas virales infecciosas (Adamson 2012). La PR se activa concomitante con la liberación del virus por fenómenos de gemación. Como la PR únicamente se encuentra activa en forma de dímero, se piensa que el autoprocésamiento se inicia por la dimerización de dos dominios de PR que están alojados en el precursor GagPol (Fun *et al.*, 2012). La maduración desencadena un segundo evento de ensamblaje que genera un núcleo cápside cónica condensada, que organiza el genoma de ARN viral y las proteínas virales para facilitar la replicación viral en la siguiente ronda de infección

(Adamson 2012). La posición y longitud de las proteínas Gag y Pol generadas durante la maduración o procesamiento del virus por la PR viral se muestran en la **Figura I.4**. Se requiere un total de 11 reacciones proteolíticas para generar un virión viable (**Figura I.4**). Cada reacción ocurre en un único sitio de procesamiento o *cleavage site* (CS) formado por 10 aminoácidos (aa) que difiere en su composición aminoacídica (de Oliveira T *et al.*, 2003; Torrecilla *et al.*, 2014). Aunque se ha descrito una velocidad procesamiento diferente de cada una de las 11 reacciones proteolíticas, el mecanismo específico que controla esa velocidad aún no se entiende totalmente (Lee *et al.*, 2012). En la **Figura I.4** se señalan en color rojo los diferentes sitios de corte o CS en el genoma del VIH-1. El nº de CS no corresponde al orden de escisión sino a la posición en los precursores Gag y GagPol.

Figura I.4. Sitios de procesamiento proteico en precursores virales Gag y GagPol del VIH-1 (A) y secuencias de aa (B)

(A)



(B)

		SITIO DE CORTE						
POLI- PROTEINAS		Nº	Nombre	Secuencia cons*	Secuencia HXB2	Posición nt (HXB2)	Posición aa (HXB2)	Nº aa en Gag y Pol
Pr160 ^{GagPol}	Pr55 ^{Gag}	CS1	P17/P24	VSQNY/PIVQN	VSQNY/PIVQN	1171-1200	391-400	127-136
		CS2	P24/P2	KARVL/AEAMS	KARVL/AEAMS	1867-1896	622-631	358-367
		CS3	P2/P7	TT-IM/MQRGN	<u>S</u> ATIM/MQRGN	1906-1935	636-646	372-381
		CS4	P7/P1	ERQAN/FLGKI	ERQAN/FLGKI	2071-2100	692-701	427-436
		CS5	P1/P6 ^{gag}	RPGNF/LQSRP	RPGNF/LQSRP	2119-2148	708-717	443-453
		CS6	P7/TFP	ERQAN/FFREN	ERQAN/FFRED <u> </u>	2071-	692-700	<u>427-</u>
		CS7	TFP/P6 ^{pol}	ENLAF/QQGEA	EDLAF/ <u>LQ</u> GKA	2097-2124	699-708	
		CS8	P6 ^{pol} /PR	VSLSF/PQITL	VS <u>F</u> NF/PQ <u>V</u> TL	2239-2267	746-755	<u>-51</u>
		CS9	PR/RT ^{p51}	CTLNF/PISPI	CTLNF/PISPI	2533-2564	845-854	151-160
		CS10	RT ^{p51} /RT ^{p66}	GAETF/YVDGA	GAETF/YVDGA	3855-3883	1285-1294	591-600
		CS11	RT ^{p66} /INT	IRKVL/FLDGI	IRKVL/FLDGI	4215-4244	1405-1414	711-720

Nº, posición CS en los precursores Gag y GagPol. Secuencia cons*: secuencia consenso de consensos sacada del GenBank (ver métodos); HXB2, secuencia de referencia del subtipo B (número de acceso en GenBank K03455); nt: nucleótido; aa: aminoácido. La longitud de cada CS es de 30 nt, por tanto, 10 aa. Los aa subrayados muestran los cambios en la secuencia HXB2 en comparación con la secuencia consenso de consensos de GenBank. CS, sitio de corte o *cleavage site*; P17, matriz; P24, cápside; P2, péptido espaciador 1; P7, nucleocápside; P1, péptido espaciador 2; TFP, proteína *transframe*; P6^{pol}, proteína P6 generada a partir del precursor Gag (Pr^{55Gag}); P6^{gag}, P6^{pol}, proteína P6 generada a partir del precursor GagPol (Pr160^{GagPol}); PR, proteasa; RT, retrotranscriptasa; IN, integrasa.

UBICACIÓN Y PAPEL BIOLÓGICO DE LAS PROTEÍNAS GAG Y POL

Las proteínas Gag, generadas a partir del procesamiento proteico de los precursores Pr⁵⁵Gag y del precursor Pr¹⁶⁰GagPol llevado a cabo por la PR viral, son esenciales para el virus, al tener papel estructural y/o funcional. Algunas desempeñan un papel biológico durante la morfogénesis, la salida del virus de la célula infectada y la infectividad viral, entre otras (**Tabla I.1**). La longitud y ubicación de proteínas Gag y Pol se muestran en la **Tabla I.2**.

Tabla I.1. Funciones de las proteínas Gag y Pol

PROTEÍNAS	REGIÓN	FUNCIÓN	DETALLES
P17 (MA)	GAG	Estructural	Formación de la matriz viral, estabiliza la envoltura. Acompaña al ADN al interior del núcleo antes de su integración.
P24 (CA)	GAG	Estructural	Formación de la cápside viral.
SP1 (P2)	GAG	Reguladora	Ensamblaje y morfogénesis, estabilidad del dímero ARN-VIH-1, e infectividad viral.
P7 (NC)	GAG	Estructural y reguladora	Formación de la nucleocápside viral, reconocimiento viral e incorporación del ARN al nuevo virión.
SP2 (P1)	GAG	Reguladora	Regulación del número de proteínas Gag y Pol.
P6	GAG	Reguladora	Responsable de incorporación de <i>vpr</i> al virión. Implicada en salida viral.
PR	POL	Funcional	Procesamiento de los precursores proteicos Gag y Gag-Pol.
RT	POL	Funcional	Media la conversión del ARN viral en el ADN proviral.
IN	POL	Funcional	Integración del ADN proviral en el genoma huésped.

P17, matriz; P24, cápside; P7, nucleocápside; P6, proteína P6 generada a partir de la poliproteína precursora Gag; PR, proteasa; RT, retrotranscriptasa; IN, integrasa.

Tabla I.2. Longitud y ubicación de las proteínas Gag y Pol generadas durante la maduración viral

VIH-1 Proteínas		Nucleótidos		Aminoácidos	
		Nº	Posición en VIH	Nº	Posición en Gag y Pol
Gag	P17	396	790-1185	132	1-132
	P24	696	1186-1881	232	133-363
	P2	39	1882-1920	13	363-377
	P7	165	1921-2085	55	378-432
	P1	48	2086-2133	16	433-448
	P6	156	2134-2292	52	449-500
	Total	1500	790-2292	500	1-500
Pol	PR	297	2253-2549	99	57-155
	RT ^{p51}	1320	2550-3869	440	156-595
	RT ^{p66}	1680	2550-4229	560	156-715
	IN	867	4230-5096	289	716-1004
	Total	2844	2253-5096	948	57-1004

Posición respecto a la secuencia de referencia HXB2 (número de acceso en GenBank K03455). Nº, número; P17, matriz; P24, cápside; P2, péptido espaciador 1; P1, péptido espaciador 2; P7, nucleocápside; P6, proteína P6 generada a partir de la poliproteína precursora Gag; PR, proteasa; RT, retrotranscriptasa con las isoformas RT^{p51} y RT^{p66}; IN, integrasa.

DIVERSIDAD GENÉTICA DEL VIH

Una de las principales características del VIH-1 es su gran capacidad de variabilidad o heterogeneidad genética debido a la combinación de tres cualidades que presenta el virus: su alta tasa de replicación, su elevada frecuencia de mutación y su alto grado de recombinación. Se calcula que se producen y destruyen entre 10^{10} y 10^{12} nuevos viriones cada 2,5 días, que es el tiempo transcurrido entre generaciones virales (Perelson *et al.*, 1996). Mientras tanto, por cada ciclo replicativo se producen, aproximadamente de 1 a 10 cambios de nt de media y de 2 a 3 eventos de recombinación en cada cadena de ARN viral por ronda de replicación (Jetzt *et al.*, 2000; Zhuang *et al.*, 2002).

Las mutaciones, incluyendo inserciones y deleciones, permanecen en el genoma debido a que la RT carece de actividad correctora de errores, siendo incapaz de eliminar los nt erróneamente incorporados durante la transcripción inversa. Todo ello produce una alta diversidad genética en cada paciente y a nivel poblacional. De modo que, dentro de un individuo infectado, el virus se organiza como una población altamente heterogénea de distintas variantes con genomas íntimamente relacionados pero no idénticos. Esta estructura poblacional, llamada “cuasiespecie viral”, se caracteriza por un proceso constante de generación de mutantes, por la competición entre ellos y por la acción de eventos de selección positiva y/o negativa en las variantes generadas. Existe un equilibrio poblacional entre los nuevos mutantes y la eliminación de los menos viables. Estos procesos de selección tienen como resultado la dominancia del genoma o genomas virales con mayor eficacia biológica o *fitness* viral (variantes mayoritarias) rodeado de un amplio espectro de mutantes (variantes minoritarias) con un extenso rango de fenotipos (Briones y Domingo, 2008).

En teoría, todas estas variantes circulantes pueden integrarse en forma de provirus en las células y así estar representadas en el reservorio de linfocitos latentemente infectados. Si el medio en el que se replican cambiase, por ejemplo por la respuesta inmunitaria o por la exposición a fármacos ARV, la selección positiva otorgaría prioridad a la variante de la cuasiespecie portadora de aquellos cambios que supusieran una ventaja competitiva en ese nuevo ambiente replicativo. La presión de selección actúa sobre la estructura poblacional del VIH pero no sobre genomas individualizados,

favoreciendo una rápida adaptación del virus a los cambios. De esta manera cada individuo VIH+ estaría infectado por partículas de VIH con genomas diferentes pero genéticamente relacionados, sometidas a procesos de selección y sucesos al azar, lo cual permite evolucionar genéticamente al virus. Desde un punto de vista evolutivo, la existencia de cuasiespecies víricas es la consecuencia del proceso continuo de mutación, competición y selección de los genomas mejor adaptados a las condiciones que rodean al virus, como puede ser la presión selectiva ejercida por el sistema inmune o la presencia de fármacos ARV.

La extraordinaria variabilidad genética en el VIH tiene implicaciones clínicas y epidemiológicas importantes (Geretti 2006; Rouet *et al.*, 2007) e implica un enorme desafío para el diseño de fármacos ARV, vacunas y terapias efectivas. Así mismo, estos viriones pueden presentar un amplio rango de fenotipos respecto a virulencia, tropismo celular, velocidad de replicación, composición antigénica y susceptibilidad a los fármacos ARV, dando al virus la capacidad de escapar al sistema inmune y al TAR.

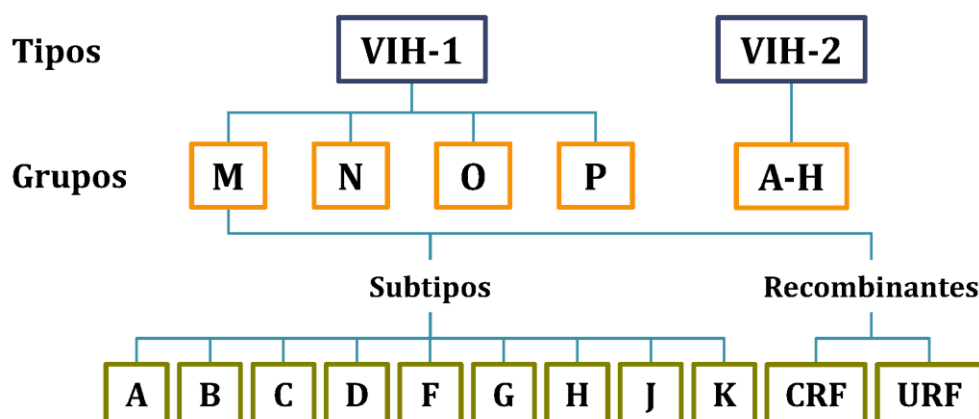
VARIANTES DEL VIH

A nivel global circulan numerosas variantes diferentes del VIH, cuya identificación y estudio se encarga la epidemiología molecular. El VIH se clasifica en dos tipos: 1 y 2 (Clavel *et al.*, 1986; Robertson *et al.*, 2000). El VIH-1 es el más extendido y es el responsable de la mayor parte de los casos de infección por VIH en el mundo. Mientras que el VIH-2, identificado en 1986 (Clavel *et al.*, 1986), es más cercano filogenéticamente al VIS que al VIH-1 y parece ser menos patogénico y menos transmisible.

Basándose en su homología genética el VIH-1 ha sido clasificado en cuatro grupos: grupo M (*main* o principal), grupo O (*outlier* o atípico), grupo N (no-M, no-O) y el grupo P. A su vez, el grupo M se ha subdividido en 9 subtipos denominados por letras (A, B, C, D, F, G, H, J, K) (Robertson *et al.*, 2000; Peeters M, 2000) y múltiples recombinantes entre ellos (**Figura I.5**). Los virus recombinantes o mosaico del VIH-1 llevan fragmentos genómicos de distintos subtipos de los viriones parentales. Se originan por fenómenos de recombinación derivados de una coinfección viral (infecciones sucesivas) o una superinfección (infecciones simultáneas) a partir de dos subtipos diferentes del virus en

un mismo hospedador. Esto ocurre en lugares donde hay una alta tasa de infección por VIH y circulan a la vez múltiples cepas del virus.

Figura I.5. Esquema de la clasificación de las variantes del VIH



Las formas recombinantes pueden ser formas recombinantes circulantes (CRF, *circulating recombinant forms*) o únicas (URF, *unique recombinant forms*). Los CRF son recombinantes entre subtipos con secuencia completa y puntos de recombinación comunes, que han sido reconocidos en más de tres individuos infectados no relacionados epidemiológicamente. Hasta el momento se han descrito 88 CRF (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>) (Cheong *et al.*, 2015; Hu *et al.*, 2017) cada uno designado por un número identificativo y por los subtipos genéticos que están presentes en su genoma.

El resto de recombinantes, conocidos como URF, son recombinantes entre subtipos que se han encontrado en individuos aislados o en grupos de personas infectadas relacionadas epidemiológicamente. Al no compartir los puntos de recombinación a lo largo del genoma entre subtipos observados en los CRF conocidos, no se pueden agrupar entre ellos. Si los URF se transmiten sucesivamente y se extienden entre la población, podrían dar lugar a nuevos CRF. De hecho, ante el incremento exponencial en el número de CRF descritos, algunos autores han propuesto agrupar aquellos CRF similares en familias de recombinantes (Zhang *et al.*, 2010), como también lo hemos hecho en esta Tesis y mostramos en el **Anexo I** (Peeters M, 2000). Dichas variantes están incrementando la prevalencia y complejidad de la pandemia, incluyendo algunos países Europeos (Yebra *et al.*, 2012).

La pandemia del grupo M VIH-1 se inició en Kinshasa alrededor de la década de 1920 y su expansión en África Central estuvo relacionada con una red de transporte activo que conectaba los principales centros de población del país con otras regiones del África Subsahariana (Junqueira y Almeida, 2016). Mientras los virus del grupo M se han expandido con carácter de pandemia, los grupos O, N y P presentan baja prevalencia y se encuentran restringidos a África Subsahariana Occidental y Central. Existen evidencias de que el paso de VIS a humano en la cuenca del río Congo ayudó a la propagación no pandémica de los grupos O, N y P del VIH-1 en Camerún y países de su entorno (Faria *et al.*, 2014).

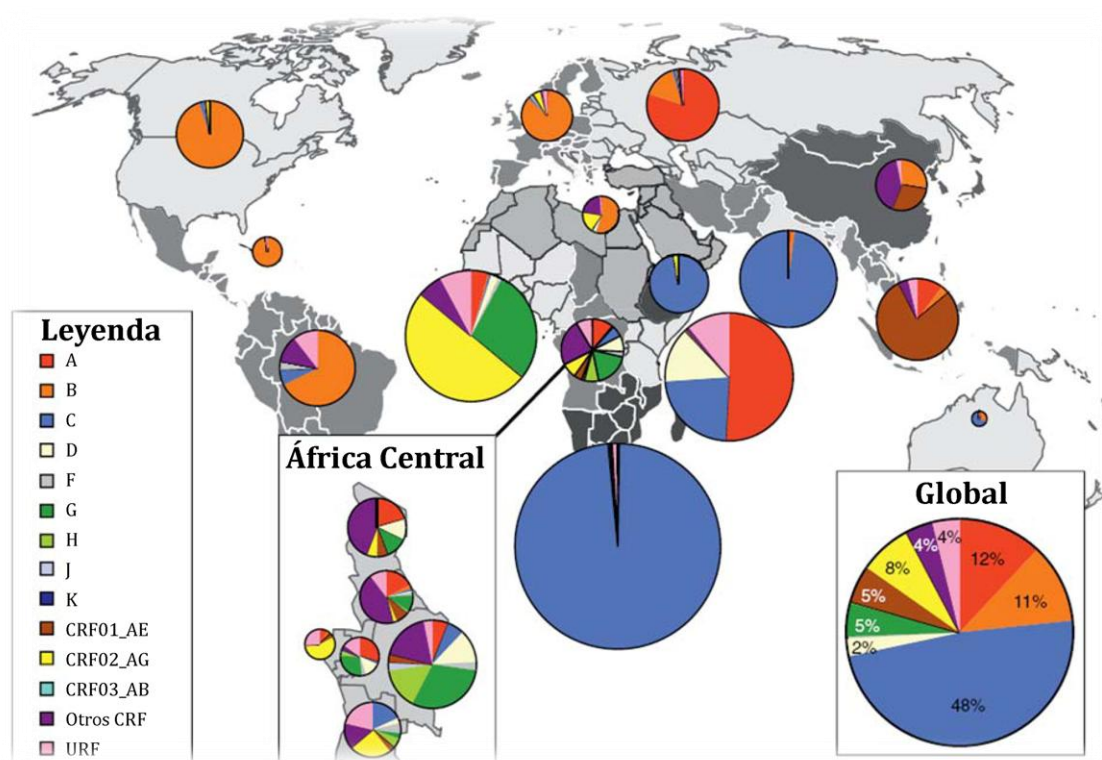
Actualmente, la distribución mundial de las distintas variantes dentro del grupo M del VIH-1 es muy heterogénea y poco a poco se comprueba como los límites geográficos entre subtipos van desapareciendo (Hemelaar *et al.*, 2011) (**Figura I.6**). El subtipo B del grupo M, que representa tan sólo el 11% de las infecciones a nivel mundial, es el predominante en Europa Occidental, Estados Unidos de América y Canadá.

A pesar de esta baja prevalencia, el subtipo B es el más estudiado y utilizado como modelo para la investigación básica y clínica del VIH-1, incluyendo el diseño y desarrollo de fármacos ARV, vacunas y test diagnósticos. Al resto de subtipos y formas recombinantes del grupo M del VIH-1 distintas al subtipo B se las denomina comúnmente “variantes no-B” del VIH-1. Estas variantes no-B son mayoritarias de manera global y son responsables de cerca del 90% de los cerca de 37 millones de infectados en el mundo.

En África Central, epicentro de la pandemia, coexisten un gran número de variantes del VIH-1. En África subsahariana, donde vive el 70% de pacientes infectados por VIH-1, se han caracterizado todos los subtipos del VIH-1 y gran variedad de formas recombinantes. La forma recombinante CRF02_AG es la más frecuente en África Occidental. En el resto del mundo la distribución del VIH es debida con elevada probabilidad a diferentes fenómenos del efecto fundador, es decir, la expansión de una variante concreta por ser la primera en llegar. De esta manera cabe destacar el subtipo C predominante en África del Sur y Oriental y también en la India; el subtipo A, en los países de la antigua Unión Soviética y en ciertas regiones de África Oriental; así como el recombinante CRF01_AE como variante predominante en el sudeste asiático. Se han

observado brotes epidémicos causados por ciertos CRF en algunas partes del mundo (Kouri *et al.*, 2015; Patiño *et al.*, 2015).

Figura I.6. Distribución mundial de los subtipos y algunas formas recombinantes del VIH-1



Adaptado de Hemelaar *et al.*, 2011. Datos 2004-2007

Una de las consecuencias de los movimientos poblacionales, debidos a intercambios comerciales, migración y turismo, ha sido la introducción de variantes no-B del VIH-1 en países desarrollados donde la variante mayoritaria es el subtipo B. De hecho, la presencia de variantes no-B es cada vez mayor, tras su introducción y posterior expansión, entre pacientes inmigrantes y autóctonos, desde zonas donde la infección por VIH-1 presenta altas tasas y donde las cepas no-B son prevalentes, como es el caso de África Subsahariana (Holguín *et al.*, 2007; Yebra *et al.*, 2009; Yebra *et al.*, 2011). La proporción de subtipos no-B y formas recombinantes del VIH-1 entre los nuevos diagnósticos varía entre países, alcanzando incluso el 50% en Francia y Bélgica (Frange *et al.*, 2008; Vercauteren *et al.*, 2009; Brennan *et al.*, 2010; Hemelaar *et al.*, 2011; von Wyl *et al.*, 2011; Ciccozzi *et al.*, 2012; Dauwe *et al.*, 2015). En España, la proporción de variantes

no-B también está aumentando considerablemente, representando el 10-15% de las infecciones por VIH. Aunque la variante no-B más frecuente en España es el CRF02_AG (Yebra *et al.*, 2012; Holguín *et al.*, 2008a; González-Alba *et al.*, 2011), se han detectado casos de infección por todos los subtipos no-B del VIH-1 y de múltiples formas recombinantes incluyendo recombinantes complejos (Holguín *et al.*, 2008b). De hecho, se ha descrito en España que una de cada 10 nuevas infecciones en adultos se debe a variantes no-B del VIH-1 y siete de cada 10 infecciones causadas por variantes no-B están originadas por cepas recombinantes (Yebra *et al.*, 2012).

Similares resultados han sido descritos recientemente en población pediátrica infectada por VIH-1 en nuestro país. Así, las tasas de infecciones por variantes no-B del VIH pasaron del 3,4% entre los niños y adolescentes diagnosticados en Madrid antes del año 2000 al 85,7% entre los diagnosticados en los últimos 5 años (2011-2015), todos ellos infectados por recombinantes del VIH-1 (Rojas P, PIDJ en prensa).

GENERALIDADES DEL TRATAMIENTO ANTIRRETROVIRAL

El ciclo infectivo del VIH muestra varios pasos clave que han sido elegidos para el diseño de fármacos ARV que los bloqueen. La RT, IN y PR, enzimas fundamentales del ciclo viral (pasos 3, 4 y 7 en la **Figura I.3**, respectivamente), han sido elegidas como dianas moleculares para las principales familias de fármacos ARV componentes de los diferentes regímenes del TAR: ITIAN, ITINAN, IP e INI, respectivamente. A estas familias les han seguido los inhibidores de la entrada, que actúan sobre el receptor o sobre el correceptor (paso 1 en la **Figura I.3**) y los inhibidores de la maduración, que inhiben las reacciones de escisión, procesamiento o *cleavage* mediadas por la PR del VIH-1, que transforman la partícula del virus inmaduro no infeccioso en un virión maduro infeccioso, como se explicó previamente.

Las pautas recomendadas para el tratamiento inicial de la infección por el VIH-1, en el momento actual, consisten en una combinación de tres fármacos que incluyan dos ITIAN asociados a un INI, o a un ITINAN, o bien a un IP potenciado (GeSIDA 2017). Los fármacos ARV autorizados en Europa, agrupados por familias, se muestran en el **Anexo II**, así como las combinaciones de ARV de inicio recomendadas por las guías nacionales (GeSIDA 2017) en el **Anexo III**.

Se recomienda la administración de TAR a todos los pacientes con infección por el VIH, con o sin sintomatología, y con independencia del nº de linfocitos T CD4+. Ello ayuda a evitar la progresión de la enfermedad y disminuir la transmisión del virus. Se consideran como excepción los pacientes que mantienen carga viral (CV) indetectable de forma mantenida sin TAR (controladores de élite). En este caso, no existe información que permita valorar el efecto beneficioso del TAR, por lo que no se puede establecer una recomendación de tratamiento (GeSIDA 2017).

Sin embargo, a pesar de contar con ARV potentes con gran capacidad para suprimir la replicación del virus y consecuentemente, de mantener la CV indetectable de la persona infectada, el VIH no se puede eliminar. El TAR es incapaz de erradicar la infección porque el VIH permanece latente en reservorios celulares y anatómicos, en los que elude tanto la respuesta inmunitaria como la acción de los fármacos. Esto implica la necesidad de mantener el tratamiento de por vida, lo que entraña un elevado riesgo de efectos adversos y también de incumplimiento terapéutico y, por consiguiente, de resistencia farmacológica. No obstante, gracias al TAR una alta proporción de pacientes VIH positivos tienen una calidad de vida prácticamente normal a pesar de los posibles efectos adversos.

EVALUACIÓN Y MONITORIZACIÓN DEL TRATAMIENTO ANTIRRETROVIRAL

Los principales parámetros a evaluar durante el seguimiento del TAR son el número y porcentaje de linfocitos T CD4+, la CV plasmática, la detección de resistencias, la concentración plasmática de fármacos ARV, el tropismo viral, el nivel de adherencia al TAR y la caracterización de la variante genética infectante.

Linfocitos T CD4

La cifra de linfocitos T CD4+ representa el principal indicador del estado inmunológico. Permite estatificar la infección por VIH, evaluar el riesgo de comorbilidad o mortalidad y la vulnerabilidad a ciertas infecciones oportunistas, así como la necesidad de profilaxis y eventual discontinuación (Asboe *et al.*, 2012; GeSIDA 2015). Por ello se debe cuantificar el nº de linfocitos CD4+ antes de iniciar el TAR y de manera periódica una vez iniciado éste, y también en pacientes controladores de élite (Kimmel *et al.*, 2005; Sayana

et al., 2011; Gale *et al.*, 2013; Gatell *et al.*, 2013). Tras el estudio START (Insight Start Study Group *et al.*, 2015) se reconoce la indicación universal de TAR a todos los pacientes, independientemente del recuento de linfocitos T CD4+.

Monitorización de la carga viral plasmática

La CV o niveles en plasma de ARN del VIH-1 expresado en copias/ml o log, es la expresión de la replicación viral y por consiguiente, de la cantidad de virus circulante. Asimismo, es el parámetro de elección para monitorizar el TAR, asegurando su éxito al discriminar entre problemas de adherencia y fracaso en el TAR (HIV/AIDS. WHO 2016). La CV ha de determinarse en la valoración inicial del paciente, antes y durante el TAR (Huang *et al.*, 2001; Asboe *et al.*, 2012; Bonner *et al.*, 2013), variando la frecuencia de determinaciones en función del estado del paciente (Kieffer *et al.*, 2004) o cuando se precisa confirmar un valor no esperado. El TAR pretende controlar a valores mínimos la replicación viral, disminuyendo de forma permanente la viremia, alcanzando unos niveles de CV indetectables. Se considera CV indetectable por debajo del límite de detección de las técnicas comercializadas, que oscila normalmente entre < 20 y < 50 copias de ARN-VIH-1/ml, o incluso menos, según la técnica empleada. Mantener viremias inferiores a 50 copias de ARN-VIH-1/ml se asocia a máxima recuperación inmunológica, reducción de la transmisión de la infección por VIH y prevención de mutaciones de resistencia (MRD) (Nettles *et al.*, 2004; Siliciano JD y Siliciano RF, 2013). Si bien en la actualidad, en nuestro entorno, más del 80% de los pacientes que inician TAR pueden alcanzar una CV <50 copias/ml a las 48 semanas (GeSIDA 2017), contar con técnicas que determinen fiablemente la CV hasta 1 copia/ml permitiría conocer el grado de replicación residual comparando la capacidad de distintos TAR para inhibir esta replicación, y finalmente aumentar las opciones de erradicar la infección por VIH. Cuando la CV se mantiene detectable (>50 copias/ml) en dos determinaciones consecutivas tras 24 semanas del inicio del TAR o bien cuando se vuelve detectable en dos determinaciones consecutivas (separadas 2-4 semanas) tras alcanzar indetectabilidad, se produce un fracaso virológico (FV), independientemente de si se producen o no selección de MRD (GeSIDA 2017). Aunque éstas son las definiciones más aceptadas de FV se pueden encontrar discordancias (EACS 2016; WHO 2016).

Concentraciones plasmáticas de fármacos antirretrovirales

Conocer la concentración de ARV en plasma permitiría monitorizar la adherencia, evaluar la respuesta terapéutica, evitar toxicidad por sobredosificación y selección de MR por niveles subóptimos, optimizar la dosis en función del perfil genotípico de la cepa infectante y prevenir interacciones entre fármacos. Sin embargo, su determinación presenta limitaciones que desaconsejan su uso en la práctica clínica habitual. Se han observado resultados discordantes de eficacia en estudios prospectivos, se desconocen los rangos terapéuticos asociados a respuesta o reducción de efectos adversos, variabilidad interindividual (por sexo, edad, peso, entre otros), no se pueden medir para ITIAN y la tecnología necesaria no siempre está disponible en los laboratorios (Wertheimer *et al.*, 2006).

Determinación del HLA-B*5701

La determinación del HLA-B*5701 reduce al 50% el riesgo de presentar una reacción de hipersensibilidad a Abacavir (ABC) por lo que se debe determinar en todos los pacientes que inicien un TAR con este fármaco. Dicha reacción de hipersensibilidad consiste en un síndrome multiorgánico presente en un 5-8% de pacientes de raza blanca que toman ABC y puede ser mortal (Hughes *et al.*, 2008).

Determinación del tropismo viral del VIH-1

El tropismo viral viene definido por el uso de los correceptores de quimiocinas CCR5 o CXCR4 presentes en células inmunitarias en el momento de entrada del VIH en la célula. Existe un único fármaco comercializado, Maraviroc (MVC), integrante de la familia de fármacos antagonistas de CCR5, con una potente actividad anti-VIH pero sin beneficio en aquellos con tropismo X4 (Sayana y Khanlou, 2009; Vandekerckhove *et al.*, 2009; Fatkenheuer *et al.*, 2010). Diferentes grupos nacionales (Poveda *et al.*, 2010) y europeos (Vandekerckhove *et al.*, 2011) recomiendan determinar el tropismo antes de iniciar el tratamiento con MVC, ya sea como rescate o en pacientes sin TAR previo en los que parezca una buena elección. El tropismo se puede determinar por métodos genotípicos que determinan la secuencia de la región V3 de la gp120, siendo la técnica de referencia actual (Poveda *et al.*, 2010; GeSIDA 2017). Asimismo, existen métodos fenotípicos, como

el *Enhanced Sensitivity Trofile Assay* (ESTA), pero son técnicas complejas, caras y menos rápidas (Svicher *et al.*, 2010; Genebat *et al.*, 2011).

Determinación de la adherencia al tratamiento

El control virológico depende de múltiples factores. Una adherencia al TAR incorrecta, o incapacidad del paciente para implicarse en la elección, inicio y cumplimiento del mismo, es la principal causa de fracaso terapéutico y, en último término, de mortalidad (Hogg *et al.*, 2002). Entre las causas de mala adherencia destacan una mala relación médico-paciente, el uso de drogas, el deterioro neurocognitivo, un bajo nivel educativo, la barrera idiomática, la falta de apoyo social y los efectos secundarios del TAR (Mills *et al.*, 2006; Grierson *et al.*, 2011; Al-Dakkak *et al.*, 2013 Nachega *et al.*, 2014).

Una vez iniciado el TAR la adherencia debe monitorizarse y reforzarse coincidiendo con las visitas clínicas para evitar niveles subterapéuticos de fármaco que faciliten la replicación del VIH y desarrollo de MRD. La mejor estrategia para evitar la mala adherencia es el uso de regímenes completos en comprimido único, ya que se asocia con menores tasas y costes de hospitalización (Cohen *et al.*, 2013). Además, ante un paciente en el que se sospeche una adherencia incorrecta es preferible iniciar TAR con pautas basadas en IP potenciado, en los que la aparición de MRD es más difícil dada su elevada barrera genética (Bangsberg *et al.*, 2007). En junio de 2008 el Grupo de estudio del Sida (GeSIDA), el Plan Nacional sobre el Sida (PNS) y la Sociedad Española de Farmacia Hospitalaria (SEFH), revisaron los factores que influyen en la adherencia y las posibles estrategias de mejora (GeSIDA/SEFH/PNS, 2008).

RESISTENCIAS A ANTIRRETROVIRALES TRANSMITIDAS Y ADQUIRIDAS. IMPLICACIÓN CLÍNICA

Debido a la estructura en cuasiespecie del virus y a la imposibilidad de la supresión completa de su replicación, tras el inicio del tratamiento el VIH puede seleccionar rápidamente variantes preexistentes portadoras de mutaciones que le confieran una reducción de susceptibilidad o incluso resistencia a los fármacos ARV. Se han descrito MRD para todos los ARV desarrollados hasta el momento, independientemente del gen

diana. Estas mutaciones alteran la forma y las propiedades bioquímicas de las proteínas virales, y así son capaces de reducir la actividad del ARV. Las variantes de VIH-1 con MRD, en general tienen menor capacidad replicativa o *fitness* (Tang y Shafer, 2012).

Los episodios de fracaso terapéutico debido a la selección de virus con MRD (también llamado fracaso virológico o FV) que el paciente VIH desarrolla durante su curso clínico, perfilan diferentes patrones mutacionales en los genes diana de fármacos ARV, que han inducido al fracaso. Así, ya se han identificado las mutaciones más relevantes asociadas a resistencia a los ARV para personas no expuestas al TAR o *naïve* (mutaciones transmitidas o MRT), descritas por la OMS (Bennett *et al.*, 2009). Las MRD de virus que infectan a pacientes con experiencia previa al TAR o tratados (mutaciones adquiridas o MRA) se incluyen en el listado de la *International AIDS Society-USA* (IAS-USA) actualizado anualmente (Wensing *et al.*, 2017) (**Tabla I.3**). La transmisión de resistencias transmitidas se ha estabilizado en Europa (Hofstra *et al.*, 2016) y en España en valores menores del 9% (Monge *et al.*, 2012; Monge *et al.*, 2014). Sin embargo, la tasa de MRT a nivel mundial depende de la región geográfica estudiada (Rhee *et al.*, 2015).

La detección de variantes virales resistentes a los fármacos puede realizarse mediante técnicas genotípicas o fenotípicas (Hirsch *et al.*, 2008). Los ensayos genotípicos son los más utilizados en la práctica clínica y detectan cambios específicos o MRD en la región genética viral que codifica las proteínas diana de los fármacos ARV (PR, RT, INT y proteínas de la envuelta o *gag*). Si se emplean técnicas de secuenciación convencional (Método de Sanger), podremos detectar aquellas MRD presentes en más del 15-20% de la población viral. Sólo con secuenciación masiva o de clones moleculares detectaremos MRD en menor porcentaje.

Las técnicas fenotípicas determinan la respuesta de la población viral mayoritaria a concentraciones crecientes de los distintos fármacos ARV y son más laboriosas. Para la determinación genotípica de resistencias a fármacos ARV existen técnicas comerciales (*ViroSeq® HIV-1 genotyping system* de Abbott Molecular) y no comerciales, pero todas amplifican el gen *pol* del VIH-1 mediante PCR y secuencian el amplicón generado, para identificar la presencia de las MRD.

La interpretación apropiada de los estudios genotípicos se desarrolla mediante algoritmos computarizados que requieren una actualización continuada y hay varios disponibles en internet: <http://sierra2.stanford.edu/sierra/servlet/JSierra>; http://engine.euresist.org/data_analysis/Default.aspx?method=viral_sequence; <http://www.geno2pheno.org/>; <http://retic-ris.net/Programa1/WP3.Resistencias.aspx>; entre otros. Aunque uno de los algoritmos de interpretación de resistencia genotípica más utilizado en la mayoría de los países es el programa HIVdb de la Universidad de Stanford (<https://hivdb.stanford.edu/hivdb/by-mutations/>), y que incluye las MRA de la IAS-USA 2017 y otras (Tabla I.3).

Tabla I.3. Listado de mutaciones de resistencia a antirretrovirales en pacientes *naïve* (A) y tratados (B) por familias de fármacos

(A)

MRT en pacientes <i>naïve</i>		
ITIAN	ITINAN	IP
M41L	L100I	L23I
K65R	K101E/P	L24I
D67N/G/E	K103N/S	D30N
T69D/ins	V106M/A	V32I
K70R/E	V179F	M46I/L
L74V/I	Y181C/I/V	I47V/A
V75M/T/A/S	Y188L/H/C	G48V/M
F77L	G190A/S/E	I50V/L
Y115F	P225H	F53L/Y
F116Y	M230L	I54V/L/M/A/T/S
Q151M		G73S/T/C/A
M184V/I		L76V
L210W		V82A/T/F/S/C/M/L
T215Y/F/I/S/C/D/V/E		N83D
K219Q/E/N/R		I84V/A/C
		I85V
		N88D/S
		L90M

(B)

MRA en pacientes tratados				
ITIAN	ITINAN	IP primarias	IP secundarias	INI
M41L	V90I	D30N	L10V/I/R/F/C	T66IAK
A62V	A98G	V32I	V11I	L74M
K65R/E/N	L100I	M46I/L	G16E	E92Q
D67N	K101E/H/P	I47V/A	K20R/M/I/T/V	E92QG
T69insert*	K103N/S	G48V	L24I	T97A
K70R/E	V106M/A/I	I50L/V	*V32I	F121Y
L74V	V108I	I54M/L	L33I/V/F	F121Y
V75I	E138A/G/K/Q/R	Q58E	E34Q	E138AK
F77L	V179D/F/T/L	T74P	M36I/L/V	G140AS
Y115F	Y181C/I/V	L76V	K43T	Y143RHC
F116Y	Y188C/L/H	V82A/F/T/S/L	*M46I/L	S147G
Q151M	G190S/A	N83D	*I47V	Q148HRK
M184V/I	H221Y	I84V	*G48V	N155H
L210W	P225H	N88S	*I50V	N155H
T215Y/F/rev	F227C	L90M	F53L/Y	R263K
K219Q/E	M230L/I		I54*L/V/*M/T/A/S/	
			D60E	
			I62V	
			L63P	
			I64L/M/V	
			H69K/R	
			A71V/I/T/L	
			G73C/S/T/A	
			*T74P	
			*L76V	
			V77I	
			V82*A/*F/*T/*S/I	
			*I84V	
			I85V	
			N88D/*S	
			L89M/V/I	
			*L90M	
			I93L/M	

MRT, mutaciones de resistencia transmitidas presentes en pacientes *naive* conforme al listado de la OMS (Bennet 2009); MRA, mutaciones de resistencia adquirida en pacientes en TAR según el listado de IAS-USA 2017 (Wensing *et al.*, 2017) y de la Universidad de Stanford (<https://hivdb.stanford.edu>; *Drug Resistance Summaries*). Con asterisco rojo (*), las MRA que no aparecen en el listado IAS-USA 2017 pero sí en la web de Stanford. En azul se indican mutaciones sólo presentes en uno de los listados. “En negrita”, mutaciones primarias. Sombreadas en gris, mutaciones que pueden ser tanto mutaciones primarias (“major mutations”) en IAS-USA 2017 como mutaciones secundarias o accesorias en Stanford (“minor mutations”). *69 inser: Complejo de inserción 69 de 2 o más amino ácidos (S-S, S-A, S-G, u otros), asociado a resistencia a todos los ITIAN aprobados actualmente por la FDA cuando se presentan con 1 o más mutaciones asociadas a análogos de timidina (TAM) en los codones 41, 210, o 215.

La transmisión de variantes del VIH-1 resistentes a ARV ha sido ampliamente documentada y asociada a fracaso al TAR (Little *et al.*, 2002; Kuritzkes *et al.*, 2008). Estos mutantes resistentes pueden quedar archivados por meses o años en el ADN proviral, en el genoma de la célula hospedadora y emerger al ser seleccionados por el TAR (Persaud *et al.*, 2007). Por ello, se debe de realizar ensayo de resistencia genotípica en todo paciente en el momento del diagnóstico y antes de empezar TAR, para confirmar si no se ha infectado con una cepa ya resistente que pudiera comprometer su primer TAR (Li *et al.*, 2011; Wittkop *et al.*, 2011; Cozzi-Lepri *et al.*, 2015). También se recomiendan estudios de resistencia en aquellos pacientes en tratamiento que presentan fracaso al TAR, al no controlar su CV. La no detección de MRD durante un FV sugiere falta de adherencia al TAR o ausencia de presión farmacológica (GeSIDA 2017). En el caso de resistencias en la IN, únicamente se recomienda su estudio si existe alta sospecha de transmisión de resistencias a esta familia (multirresistencia en PR y/o RT o caso índice tratado con INI) (GeSIDA 2017).

VARIABILIDAD NATURAL EN RESIDUOS DE POL RELACIONADOS CON RESISTENCIA A ANTIRRETROVIRALES

Existen polimorfismos presentes de manera natural a lo largo del genoma, conocidos como polimorfismos naturales, específicos de uno o varios subtipos o recombinantes y fijados en el genoma durante la evolución viral. Algunos de esos polimorfismos se localizan en algunas posiciones de las proteínas de ciertas variantes del VIH relacionadas con resistencia a ARV que pueden afectar a la susceptibilidad de fármacos ARV (**Tabla I.4**). Así, se ha visto que polimorfismos naturales en VIH-2 son responsables de la resistencia natural a los ARV ITINAN, a algunos IP y a enfuvirtida (ENF) (Parkin *et al.*, 2004; Witvrouw *et al.*, 2004). También se han observado en ciertas variantes no-B del VIH-1 y en los subtipos B (Yebra *et al.*, 2010a; Yebra y Holguín 2010b; Llácer Delicado 2016).

Tabla I.4. Polimorfismos y resistencias naturales en variantes no-B del VIH-1

	Polimorfismos naturales en variantes no-B del VIH-1*	Impacto	Mutaciones de resistencia
ITIAN	Ninguno	Sensible a ITIAN	Mismas que en subtipo B Subtipo C: mayor frecuencia de K65R Subtipo A: menor frecuencia de K65R que en subtipos B y C Sugeridas vías de adquisición de TAM diferentes entre variantes CRF06_cpx parece desarrollar TAMS más rápido que CRF02_AG
ITINAN	Grupo O: A98G, V179E, Y181C	Resistencia natural del Grupo O a ITINAN	Subtipo C: se adquiere antes V106M que en B. Mayor frecuencia de K103N e Y181C que subtipos A y D y sobreestimación de resistencia a rilpivirina (RPV) y a etravirina (ETV) por algunos algoritmos. Subtipo F: mayor frecuencia de V106I que en subtipo B
IP	Grupo O: Q58E Variantes no-B: L10VI, V11I, G16E, K20RI, L33IV, M36ILV, D60E, I62V, L63P, I64MV, H69KR, A71V, V77I, V82I, L89MI, I93L (mutaciones secundarias)	Sensible a IP** Algunos de estos cambios tienen efecto en la barrera genética de IPs o pueden promover vías de resistencia alternativas	Mismas que en subtipo B Subtipo C: resistencia a nelfinavir (NFV) vía L90M vs. vía D30N (subtipo B) L89MI promueve aparición de L90M K20I y M36I mejoran capacidad replicativa con IP <i>in vitro</i> No se recomienda utilizar algoritmo ANRS en cepas con polimorfismos naturales en posiciones M36, H69 y L89 para la interpretación de resistencias a tipranavir (TPV)
IF	Grupo O y N: S138A	No se conoce el impacto	Mismas que en subtipo B
INI***	E157Q (CRF03_AB) G163R (CRF44_BF)	No se conoce el impacto	Mismas que en subtipo B Subtipo C: N155H confiere menor nivel de resistencia
MVC		Sensible a MVC	Determinación del Tropismo: Existe una mayor variabilidad de la secuencia V3. Los algoritmos genotípicos están optimizados para el subtipo B del VIH-1. Interpretación igual que para subtipos B excepto para variantes de subtipo C, en las que se recomienda usar la "matriz C" de PSSM

Tabla elaborada y cedida por la Dra. África Holguín de las próximas Guías de resistencias de la Red Española de Investigación en Sida (RIS) 2017-2018.

* Polimorfismos presentes en >90% de secuencias virales en pacientes naïve de al menos alguna variante del VIH-1, incluyendo grupos O, N, P; subtipos distintos del B (A, C, D, F, G, H, J, K) y los 79 recombinantes circulantes (CRF) del grupo M del VIH-1 descritos hasta la fecha de elaboración de la tabla (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>).

** Se requieren más estudios para entender mejor la susceptibilidad a cada IP de las variantes con polimorfismos naturales en una o más posiciones de resistencia secundaria y su efecto en el desarrollo de resistencias primarias.

*** Tres cambios en la integrasa previamente asociados a resistencia a INI (V72I, L101I, T124A), se podrían considerar polimorfismos naturales en ausencia de terapia en un gran número de variantes diferentes del subtipo B del VIH-1 (Llácer Delicado 2016).

ITIAN: inhibidores de la transcriptasa inversa análogos de nucleos(t)ido; ITINAN: inhibidores de la transcriptasa inversa no análogos de nucleósido; IP: inhibidores de la proteasa; INI: inhibidores de la integrasa; MVC: maraviroc.

Estos polimorfismos naturales asociados a resistencia a ARV y presentes en ciertas variantes del VIH (tipos, grupos, subtipos, CRF) podrían comprometer la eficacia terapéutica de estos fármacos en los pacientes infectados con esas cepas. Así, algunos polimorfismos naturales en posiciones de la PR relacionadas con resistencia a IP podrían conferir una gran susceptibilidad a IP en subtipos no-B del VIH-1 (González *et al.*, 2003; Santos *et al.*, 2012), mejorar la capacidad replicativa viral *in vitro* (Holguín *et al.*, 2006b), reducir la barrera genética (número de mutaciones que el virus necesita acumular para desarrollar resistencia) de algunos fármacos IP (Van de Vijver *et al.*, 2005) o afectar a la afinidad de unión entre fármaco y sustrato (Kinomoto *et al.*, 2005). Ello refuerza la necesidad de tomar en cuenta las variantes del VIH-1 en la interpretación de los estudios de resistencias.

GAG COMO DIANA TERAPEÚTICA

A pesar de que los diferentes regímenes de TAR disponibles en la práctica clínica se basan en la inhibición de proteínas virales (PR, RT e IN), o de la unión de gp120 con los correceptores celulares (CCR5), o de la fusión virus-célula, hay otras proteínas virales que son dianas potenciales de fármacos. Las proteínas Gag son una nueva diana ARV potencial, se están desarrollando nuevos compuestos que interfieren con la generación de proteínas Gag activas y, por tanto, con la maduración del virus (Adamson *et al.*, 2009). Bevirimat fue el primer candidato a los inhibidores de maduración, al impedir la formación de proteína P24 de la CA viral, inhibiendo el procesamiento de los precursores virales Gag y Gag-Pol por la PR viral (Li *et al.*, 2006). Aunque podría ser activo en pacientes infectados por distintas variantes del VIH-1 (Yebra y Holguín, 2008), ha presentado limitaciones y actualmente hay más compuestos de esta familia de fármacos ARV inhibidores de maduración en desarrollo clínico (Blair *et al.*, 2009; Titolo 2010; Urano *et al.*, 2015; Spearman 2016).

MUTACIONES EN GAG IMPLICADAS EN RESISTENCIAS A INHIBIDORES DE LA PROTEASA

Se ha observado una asociación entre mutaciones en ciertas proteínas Gag y mutaciones de resistencia a inhibidores de la proteasa (MRD-IP) en pacientes en FV en

tratamiento con IP, incluso en ausencia de mutaciones en la PR (Fun *et al.*, 2011). Estos cambios se pueden ubicar en ciertos sitios de procesamiento proteico para generar las proteínas Gag (*Gag Cleavage Sites* o GCS) (Nijhuis *et al.*, 2007; Sutherland *et al.*, 2015a). Por ello, cambios en Gag han sido propuestos como un nuevo mecanismo de resistencia a IP, que explicaría dicha resistencia a IP en ausencia de MRD en la PR (Nijhuis *et al.*, 2007; Sutherland *et al.*, 2015a; Giandhari *et al.*, 2016). Sin embargo, ciertas mutaciones en el gen *gag* contribuyen a resistencia a IP, pero sólo en presencia de MRD primarias a IP (Giandhari *et al.*, 2016).

Estos cambios en Gag dentro o fuera de los GCS pueden restaurar defectos en la capacidad replicativa o *fitness* del VIH-1, asociados a mutaciones en la PR en virus resistentes a IP, actuando como mutaciones compensadoras (Adamson *et al.*, 2009; Kaufmann *et al.*, 2001). También pueden aumentar la eficacia del corte por la PR salvaje (*wild type*) conduciendo a la resistencia a IP. Se ha observado que virus con PR resistentes, que alteraban la capacidad proteolítica del corte en el GCS P7/P1, podrían seleccionar cambios en *gag* que recuperan la capacidad replicativa del virus y permitir el procesamiento óptimo en ese sitio por la PR mutante (Van Maarseveen *et al.*, 2012). Por tanto, mutaciones en *gag* contribuyen a la resistencia a IP en pacientes tratados compensando la pérdida de *fitness* viral y se ha sugerido que la región amino terminal de *gag* influiría en la susceptibilidad a IP y capacidad replicativa (Parry *et al.*, 2011). La coevolución *gag* y *pol in vivo* podría ser empleada por el virus para evadir la presión selectiva realizada por los fármacos ARV y podría esperarse una asociación entre patrones mutacionales en *gag* con ciertas MRD a fármacos surgidas tras el TAR y especialmente tras el fracaso terapéutico, sobre todo en TAR con IP (Banke *et al.*, 2009; Knops *et al.*, 2010; Fun *et al.*, 2012). La coevolución de la PR y su sustrato *gag* durante el tratamiento con IP podría afectar potencialmente al tratamiento con futuros inhibidores de maduración (Fun *et al.*, 2011).

Existe una asociación entre MRD a ciertos inhibidores de la maduración y MRD-IP (Verheyen *et al.*, 2010). Así, se ha encontrado en análisis longitudinales de pacientes en FV en tratamiento de primera línea con IP, incluyendo lopinavir (LPV), que la mutación 76V en la PR, previamente asociada a fracaso con LPV, se asociaba con los cambios M46I en la PR y A431V en el sitio de procesamiento NC/P1 de Gag (Knops *et al.*, 2010). Por

tanto, este sería un caso donde cambios en *gag* cooperan con mutaciones en PR originando resistencia a IP. Este cambio en *gag* ya había sido descrito previamente en el 23% de pacientes con FV a IP (Bally *et al.*, 2000).

El polimorfismo natural P453L en *gag* se relacionó con MRD-IP vía I84V *versus* V82I (Bally *et al.*, 2000). Otros autores sugirieron distintas asociaciones entre ciertos cambios en *gag* con MRD-IP: A431V (en NC) asociado a I54V y V82A en la PR, I437V (en P1) con I54V-V82F/T/S de la PR y L449F/R452S/P453L (en P6) con D30N-I84V en PR (Verheyen *et al.*, 2006). Se ha descrito que virus con L76V en la PR o I437T/V en *gag* podrían inducir resistencia al IP darunavir (DRV) sin necesidad de existir ninguna otra mutación (Lambert-Niclot *et al.*, 2008; Larrouy *et al.*, 2011). A pesar de lo previamente expuesto, hay pocos estudios genéticos sobre variabilidad natural de *gag* y de los GCS y de su relación con resistencia a fármacos y fracasos terapéuticos para su aplicación clínica.

Por ello, esta Tesis se dedica principalmente a estudiar la variabilidad natural que existe en los codones asociados a resistencia en la PR, RT, IN y Gag del VIH-1, así como en los sitios de procesamiento proteico que originan las proteínas virales Gag y Pol, en todas las variantes del VIH-1. Para ello hemos incluido el panel más amplio de variantes del VIH-1 analizado hasta la fecha, incluyendo por primera vez secuencias de los 4 grupos del VIH-1 y de todos los subtipos y todos los recombinantes CRFs con secuencia disponible en el GenBank en el momento de estudio.

MATERIAL Y METODOS

“Si puedes encontrar un camino sin obstáculos, probablemente no te lleve a ninguna parte”

Frank A. Clark (1857-1945)

SECUENCIAS DEL VIH-1

Para realizar esta Tesis se emplearon dos fuentes de secuencias del virus, en función de cuales fueran los objetivos planteados:

1. Secuencias de distintas variantes del VIH-1 en diferentes regiones genéticas del virus (*gag*, *pol* CDS, PR, RT e IN) disponibles en la base de datos del Laboratorio Nacional de los Álamos de EE. UU. (LANL), (<https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html>) que recoge automáticamente todas las secuencias del VIH unas semanas después de que hayan sido depositadas en el GenBank (<http://www.hiv.lanl.gov/content/sequence/GENBANK/genbank.html>). Dado que es el origen de las secuencias, emplearemos el término GenBank como fuente de secuencias cuando corresponda a lo largo de esta Tesis.
2. Secuencias de VIH-1 recuperadas de muestras clínicas de pacientes infectados disponibles en el Laboratorio de Epidemiología Molecular del VIH (EpiMolVIH), el cual está adscrito al Departamento de Microbiología y Parasitología del Hospital Universitario Ramón y Cajal (RyC) y al Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS).

DESCARGA DE SECUENCIAS DE VIH DE LA BASE DE DATOS GENBANK

En el GenBank se recopilan todas las secuencias genéticas del VIH-1 y VIH-2, dándoles un número de acceso para su identificación y posibilidad de descarga desde cualquier ordenador. Fue creado por la División de SIDA del Instituto Nacional de Alergia y Enfermedades Infecciosas (NIAID), que forma parte del *National Institutes of Health* (NIH) y constituye una colección de disponibilidad pública de secuencias de ADN que se actualiza periódicamente.

La página web también permite el acceso a un gran número de herramientas bioinformáticas que pueden ser utilizadas para la descarga y análisis de secuencias.

Para nuestro estudio descargamos de GenBank todas las secuencias pertenecientes a cada una de las variantes del VIH-1 existentes en formato fasta, que es un fichero en formato texto utilizado para representar secuencias de ácidos nucleicos o de proteínas requeridas para análisis bioinformáticos. En caso de la descarga de secuencias de nt, éstos se pueden traducir a aa si se requiere para el análisis. En el formato fasta los pares de bases nucleotídicos (A, adenosina; C, citosina; G, guanina; T, timidina) o los aa se representan usando códigos de una única letra, como se indica en la sección de abreviaturas. La **Figura M.1** indica cómo se selecciona desde GenBank la región genética cuyas secuencias queremos descargar de la variante o variantes del VIH-1 de interés.

Figura M.1. Descarga de secuencias de la Base de Datos del GenBank

HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search Site

Sequence Search Interface

Accession number
Sequence name
Sequence length
exact ☒ Sampling year
Sampling country

Virus HIV-1
Subtype Any subtype
No subtype
A
A1
A2
B

☐ Include recombinants

More sequence information

Find all sequences for a specific gene or region (HIV-1, SIVcpz and SIVgor)

Genomic region Gag-Pol
Gag
p17 (matrix)
p24 (capsid)
p7 (nucleocapsid)
p6
Pol CDS

Or define start and end
☐ Include fragments of minimum length 100

- **A** Selección de la variante VIH-1 a descargar.
- **B** Selección de la región genómica de interés.

La **Figura M.2** muestra un ejemplo de secuencia fasta de la PR viral. En el formato fasta, la línea de cabecera, que comienza con '>', proporciona un nombre y/o un identificador único a la secuencia, de la región viral y del número de aa y nt, junto con otra información adicional si procede (tipo de paciente, estudio, etc). La simplicidad del formato fasta hace fácil el manipular y analizar secuencias.

Figura M.2. Secuencia de la PR del VIH-1 en formato fasta

```
>RDC_P3 PR (99aa, 297 nts)
CCTCAAATCACTCTTTGGCAACGACCCGTTGTCACAGTAAACATAGAGGGACAGATAAGGGAAGCTCTATTAGATACAG
GAGCAGATGATACAGTATTAGAAGAAATAAATTTGTCAGGAAGATGGAAGCCAAAAATGATAGGGGGAATTGGAGGTTT
TATCAAAGTAAGACAGTATGAGCAAGTAACCTTAGAAATCTGTGGGAAAAAGGCTATAGGTACAGTATTAGTAGGGCCT
ACACCTGTCAATATAATTGGAAGGAATATATTGACTCAAATTGGTTGCACTTTAAATTTT
```

De esta manera recopilamos secuencias pertenecientes a de cada uno de los 4 grupos (M, O, N, P), 9 subtipos del grupo M (A, B, C, D, F, G, H, J y K), sub-subtipos (A1, A2, F1 y F2) y de todos los recombinantes CRF del grupo M del VIH-1 disponibles en el GenBank a fecha de estudio. Como secuencia de referencia empleamos la secuencia del aislado HXB2 del subtipo B del VIH-1 (nº de acceso GenBank K03455), en la secuencia genética de interés.

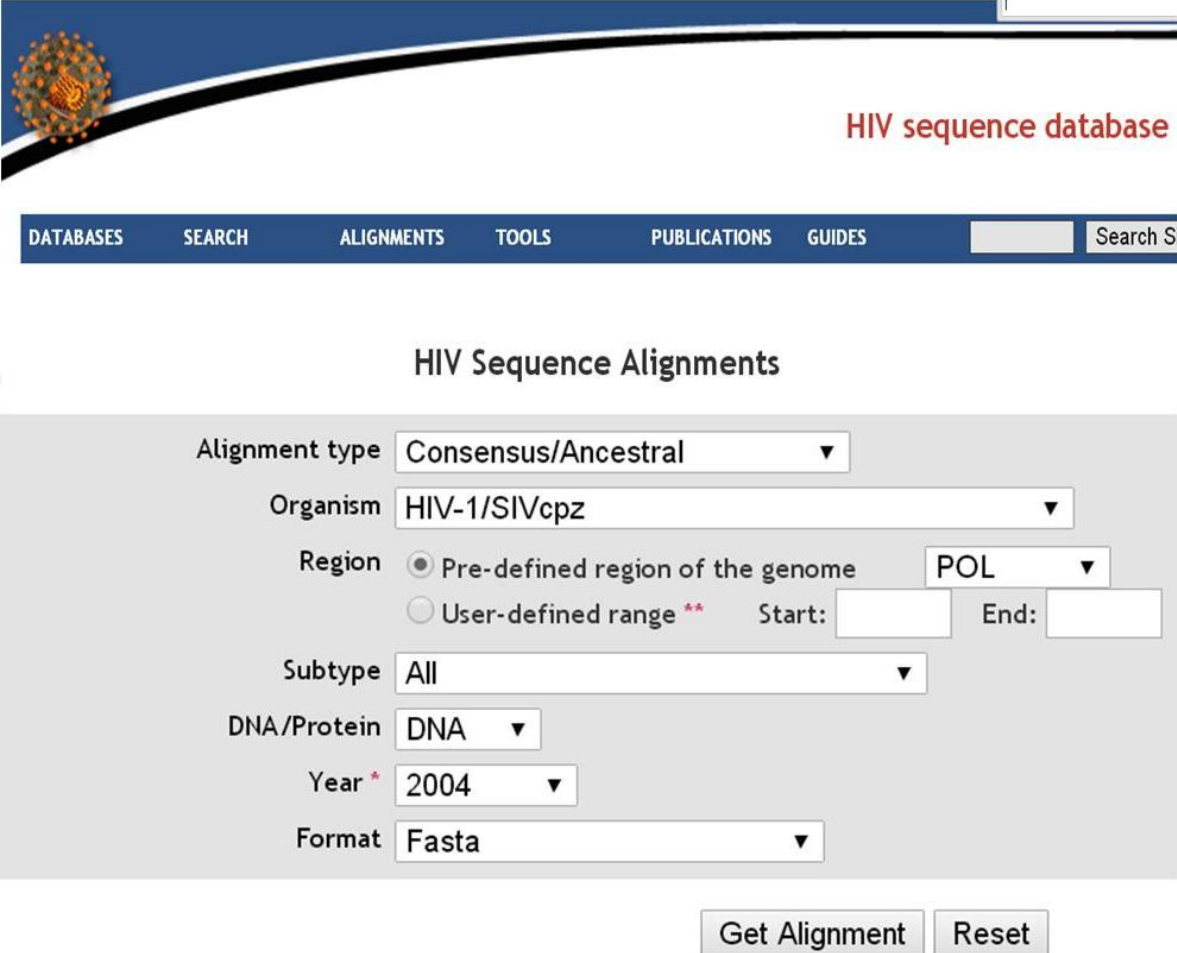
Durante esta Tesis analizamos las diferentes regiones genómicas del virus de interés para nuestro estudio: la región codificante de *gag* (que codifica para las proteínas Gag: P17, P24, P2, P7, P1 y P6 del VIH-1), de *pol* completo (Pol CDS, que codifica para las proteínas: PR, RT e IN), o de las zonas codificantes de las proteínas Pol por separado (PR, RT^{p51} e IN^{p31}). La longitud y ubicación de las proteínas Gag y Pol estudiadas se indican en la **Tabla I.1** de la Introducción. En cada capítulo se describirán las secuencias analizadas, el criterio de inclusión, la fecha de descarga, los filtros empleados, y otros procedimientos más específicos de los estudios llevados a cabo.

DESCARGA DE SECUENCIAS CONSENSO DE CADA SUBTIPO, SUB-SUBTIPO Y RECOMBINANTE CRF DEL VIH-1

Las secuencias consenso se definen como aquellas secuencias que representan a la población viral del individuo infectado. Dicha secuencia consenso tiene en cada posición nucleotídica (o aminoacídica), el nt (o el aa) más frecuente de la población viral que infecta al paciente tras realizar un alineamiento de múltiples secuencias. Generalmente, ese nt (o aa) más frecuente está presente en más del 50% de los viriones del compartimento estudiado (sangre periférica normalmente). Las secuencias consenso para las regiones del genoma *gag* y *pol* se pueden encontrar y descargar de LANL (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html#consensus>) para todos los subtipos del grupo M (A, B, C, D, F, G, K incluyendo los sub-subtipos A1, A2, F1 y F2) y para 11 de las 88 formas recombinantes circulantes descritas en la actualidad: CRF01_AE, CRF02_AG, CRF03_AB, CRF04_CPX, CRF06_CPX, CRF07_BC, CRF08_BC, CRF10_CD, CRF11_CPX, CRF12_BF, CRF14_BG. El subtipo K y la CRF07_BC no disponen de secuencia consenso para la región *pol*, sólo para *gag*; y el sub-subtipo F2 no dispone de secuencia consenso para la región *gag*, únicamente para *pol*. Asimismo, podemos encontrar una secuencia consenso del grupo M, la cual se trata de un consenso realizado a partir de las secuencias consenso disponibles para los subtipos A, B, C, D, F, G y H, obtenidas tras alinear todas las secuencias disponibles y depositadas en la base de datos de GenBank de la variante concreta en *gag* y *pol*, incluyendo sólo 1 secuencia por paciente y eliminando las secuencias muy similares y aquellas problemáticas. A esta secuencia la llamamos secuencia consenso de consensos.

Con el fin de buscar los dominios conservados en *gag* y *pol* entre variantes, seleccionamos y descargamos todas las secuencias consenso de la región genómica de interés disponibles en GenBank tras acceder a <http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>) y realizar en el campo “Options” las siguientes elecciones: “Alignment type: Consensus/Ancestral Sequences”; “Region: GAG o POL”; “Subtypes: All” y dentro de “DNA (para nt) o Protein (para aminoácidos)” (Figura M.3).

Figura M.3. Descarga de las secuencias consenso del VIH-1 disponibles en GenBank



The screenshot shows the 'HIV sequence database' interface. At the top, there is a navigation bar with links: DATABASES, SEARCH, ALIGNMENTS, TOOLS, PUBLICATIONS, GUIDES, and a search box. Below this is the 'HIV Sequence Alignments' section. The form contains the following fields and options:

- Alignment type:** Consensus/Ancestral (dropdown)
- Organism:** HIV-1/SIVcpz (dropdown)
- Region:**
 - ☒ Pre-defined region of the genome (selected)
 - ☐ User-defined range ** (with Start and End input boxes)
- Subtype:** All (dropdown)
- DNA/Protein:** DNA (dropdown)
- Year *:** 2004 (dropdown)
- Format:** Fasta (dropdown)

At the bottom right of the form are two buttons: 'Get Alignment' and 'Reset'.

GENERACIÓN DE SECUENCIAS CONSENSO A PARTIR DEL ALINEAMIENTO DE SECUENCIAS DISPONIBLES EN GENBANK

Para ciertas variantes del VIH-1, sobre todo los nuevos recombinantes CRF, no existían en GenBank secuencias definidas como consenso o mayoritarias. Por ello, procedimos a alinear todas las secuencias disponibles tanto para las variantes que ya disponían de secuencia consenso en GenBank como para todas las demás que no la tenían. Después analizamos la conservación en las distintas posiciones nucleotídicas para definir nuestras secuencias consenso propias para cada variante del VIH-1 en las que en cada posición del alineamiento se localiza el nt más prevalente en la población viral.

ALINEAMIENTO Y EDITADO DE SECUENCIAS DE NUCLEÓTIDOS Y AMINOÁCIDOS

Para alinear y editar todas las secuencias de nt recuperadas de GenBank en formato fasta, incluida la secuencia consenso de consensos descargada de GenBank, empleamos las versiones 5.0 y 6.0. del programa MEGA (*Molecular Evolutionary Genetics Analysis*, <http://www.megasoftware.net/>) (Tamura *et al.*, 2011; Tamura *et al.*, 2013), generando un alineamiento similar al que se observa en la **Figura M.4**. El programa MEGA nos permite realizar análisis moleculares relacionados con la filogenia, puesto que proporciona herramientas para buscar y analizar secuencias de proteínas o nt con perspectivas evolutivas. Este software se desarrolló para análisis comparativos de secuencias de ADN y proteínas que están dirigidas a inferir los patrones evolutivos moleculares de genes, genomas y especies con el tiempo (Kumar *et al.*, 1994; Tamura *et al.*, 2011).

Figura M.4. Alineamientos editados con MEGA en nt (A) o en aa (B)

(A)

Species/Abbrv	Sequence
1. 35 AD.AF.2006.077H.GQ477442	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT
2. 35 AD.AF.2006.047H.GQ477443	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT
3. 35 AD.AF.2006.051H.GQ477444	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT
4. 35 AD.AF.2006.073H.GQ477445	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT
5. 35 AD.AF.2007.169H.GQ477446	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT
6. 35 AD.AF.2007.273H.GQ477448	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT
7. 35 AD.AF.2007.221H.GQ477449	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT
8. 35 AD.AF.2006.007H.GQ477450	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT
9. 35 AD.AF.2006.027H.GQ477451	ATGGGTTCGAGAGCGTCAATATTAGTGGAGGAAAAATT

(B)

Species/Abbrv	Sequence
1. 35 AD.AF.2006.077H.GQ477442	MGARASILSGGLDAEIKILRGGKKKYALHIVAS
2. 35 AD.AF.2006.047H.GQ477443	MGARASILSGGLDAEIKILRGGKKKYALHIVAS
3. 35 AD.AF.2006.051H.GQ477444	MGARASVLSGGLDAEIKILRGGKKKYALHIVAS
4. 35 AD.AF.2006.073H.GQ477445	MGARASILSGGLDAEIKILRGGKKKYALHIVAS
5. 35 AD.AF.2007.169H.GQ477446	MGARASILSGGLDAEIKILRGGKKKYALHIVAS
6. 35 AD.AF.2007.273H.GQ477448	MGARASVLSGGLDAEIKILRGGKKKYALHIVAS
7. 35 AD.AF.2007.221H.GQ477449	MGARASILSGGLDAEIKILRGGKKKYALHIVAS
8. 35 AD.AF.2006.007H.GQ477450	MGARASILSGGLDAEIKILRGGKKKYALHIVAS
9. 35 AD.AF.2006.027H.GQ477451	MGARASVLSGGLDAEIKILRGGKKKYALHIVAS

Tras obtener el alineamiento procedimos a editarlo eliminando los *gaps* o espacios vacíos (sin nt o aa) dentro de las secuencias, para su posterior traducción que nos permitió convertir los alineamientos de nt en aa empleando el mismo programa MEGA.

IDENTIFICACIÓN DE LOS RESIDUOS Y DOMINIOS CONSERVADOS ENTRE VARIANTES

Una vez alineadas todas las secuencias procedentes de GenBank para cada una de las variantes VIH-1 existentes en la región o regiones genéticas virales de interés, pasamos a analizar su grado de conservación. El objetivo fue definir nuestras propias secuencias consenso en todos los grupos del VIH-1 y en todos los subtipos y recombinantes CRF del VIH-1 e identificar los residuos y dominios conservados entre variantes. También queríamos estudiar la variabilidad natural en cada residuo relacionado con resistencia a fármacos ARV asociada a subtipo y CRF específico del VIH-1. El grado de conservación se definió como el nº de residuos conservados con respecto a la secuencia consenso de consensos procedente de GenBank dividido entre el total de residuos analizados. Así estimamos el grado de conservación a nivel de aa en todas las secuencias descargadas para cada variante del VIH-1 con respecto a la secuencia consenso de consensos procedente de GenBank a lo largo de toda la secuencia, pero fijándonos en especial en la región o residuos de interés. Establecimos un código de colores para indicar diferentes grados o niveles de conservación o de variabilidad, que definiremos en cada capítulo correspondiente.

IDENTIFICACIÓN DE POSICIONES EN GAG Y POL IMPLICADAS EN RESISTENCIA A FÁRMACOS ANTIRRETROVIRALES

En el alineamiento generado, a nivel de aa de todas las secuencias disponibles en el GenBank para cada variante del VIH-1 específica con respecto a la secuencia consenso de consensos, identificamos las posiciones en las proteínas Gag y Pol (PR, RT e IN) relacionadas con la resistencia a fármacos ARV, definidas según la lista de la International AIDS Society-USA (IAS-USA) (Wensing *et al.*, 2017) y otras fuentes bibliográficas.

PREDICCIÓN DE LA SUSCEPTIBILIDAD A ANTIRRETROVIRALES

Utilizando las secuencias *pol* VIH-1 descargadas de GenBank analizamos la susceptibilidad prevista a fármacos ARV frente a *pol*, de cada secuencia viral empleando la Base de Datos de Stanford (<https://hivdb.stanford.edu/>). Para ello se introdujo el archivo con las secuencias en formato Fasta o texto (txt) dentro de la web <https://hivdb.stanford.edu/hivdb/by-sequences/> del programa HIVdb de la Universidad de Stanford (Palo Alto, California, EE. UU.). Este programa alinea automáticamente la secuencia del VIH-1 a estudiar en la PR, RT e IN frente a la secuencia de referencia del subtipo B (secuencia HXB2, nº de acceso GenBank K03455). Este algoritmo de interpretación de la resistencia genotípica HIVdb de Stanford (<http://sierra2.stanford.edu/sierra/servlet/JSierra>) clasifica la susceptibilidad o sensibilidad a 22 fármacos ARV aprobados por la Agencia Estadounidense de alimentos y fármacos (*US Food and Drug Administration*, FDA) (8 IP, 7 ITIAN, 4 ITINAN y 3 INI) en cinco categorías, dependiendo de una puntuación de penalización (*penalty score*) asignada a cada mutación de resistencia. Así, según su impacto en la resistencia a ARV los grupos de clasificación son: susceptible (*score* 0-9), nivel de resistencia potencialmente bajo (*score* 10-14), nivel bajo de resistencia (*score* 15-29), resistencia intermedia (*score* 30-59) y nivel de resistencia alto (*score* \geq 60) (<http://hivdb.stanford.edu/DR/asi/releaseNotes/index.html>).

ANÁLISIS ESTADÍSTICO

Los análisis estadísticos se realizaron utilizando el programa Epi Info v6.0 (*Center for Disease Control and Prevention*, CDC, Atlanta, EE.UU.). La significación estadística se fijó en $p < 0,05$ (<http://www.cdc.gov/epiinfo/user-guide/statcalc/introduction.html>).

CAPÍTULO 1

Variabilidad en los sitios de procesamiento proteico entre grupos, subtipos y recombinantes del VIH-1

“Nada que un hombre haga lo envilece más que el permitirse caer tan bajo como para odiar a alguien”

Martin Luther King (1929-1968)

JUSTIFICACIÓN Y OBJETIVOS

Resulta esencial conocer la variabilidad natural asociada a los distintos grupos, subtipos y recombinantes circulantes (CRF) del VIH-1, sobre todo en aquellas posiciones de las proteínas Gag y Pol virales donde aparecen los cambios de aminoácidos (aa) asociados a resistencia a fármacos antirretrovirales (ARV) que pueden conducir a un fracaso al tratamiento antirretroviral (TAR). Además, dada la alta relevancia biológica de los sitios de procesamiento proteico o *cleavage sites* (CS) durante la maduración del VIH-1 y de la influencia de ciertos cambios de aa de los CS en la recuperación de eficacia biológica o *fitness* de los virus resistentes a ARV, es necesario conocer su conservación entre las variantes del VIH-1. Ello podría ayudar a un mejor diseño de nuevos fármacos ARV de la reciente familia de los inhibidores de la maduración viral en desarrollo, y al diseño de futuras vacunas dirigidas a los CS.

Sin embargo, existen pocos datos en la literatura científica disponibles acerca de la variabilidad en los CS asociada a cada una de las variantes del VIH-1. Además, los pocos estudios que hay, han analizado un número (nº) pequeño de secuencias y de variantes distintas del VIH-1, incluyendo sólo unos pocos recombinantes del VIH-1 de los que circulan en la pandemia (de Oliveira *et al.*, 2003; Liégeois F *et al.*, 2013; Luo M *et al.*, 2013).

Por lo tanto, el objetivo de este estudio descriptivo fue analizar, por primera vez, el grado de conservación de cada CS en Gag y Pol a nivel de aa en los 4 grupos de VIH-1, en los 9 subtipos y en todos los recombinantes circulantes (CRF) del grupo M del VIH-1 que tuvieran secuencias disponibles depositadas en GenBank a la fecha de estudio (enero 2012). Además, se definieron por primera vez las secuencias consenso en cada CS en todas las variantes de VIH-1, incluyendo aquellas para las que el GenBank no las tenía disponibles. Por último, se identificaron los residuos de aa altamente conservados en cada sitio de procesamiento proteico viral o CS.

MATERIAL Y MÉTODOS

DESCARGA Y ALINEAMIENTO DE SECUENCIAS

Para este estudio la descarga, alineamiento y traducción de secuencias de nucleótidos (nt) en formato fasta codificantes para *gag/pol* se realizó, según lo descrito en el apartado de Materiales y Métodos de esta Tesis. A fecha de estudio, enero 2012, contábamos con secuencias disponibles para 51 de los 58 CRF descritos en el momento del estudio.

IDENTIFICACIÓN DE LOS SITIOS DE PROCESAMIENTO PROTEICO VIRAL O CS

Tras alinear convenientemente las secuencias descargadas, pasamos a localizar los 110 aa implicados en los 11 sitios de procesamiento proteico o CS en las poliproteínas precursoras Gag y GagPol, incluyendo 10 aa por sitio. Sólo recordar que la PR viral corta en esos CS para generar las proteínas Gag (P17 o MA, P24 o CA, P7 o NC, P6 y dos espaciadores peptídicos -P1 y P2-) y las proteínas Pol (PR, RT e IN) virales, como se explicó previamente en la Introducción de esta Tesis. Asimismo, la ubicación de los 11 CS en el genoma del VIH-1 y su secuencia en aa en el aislado de referencia HXB2 del subtipo B del VIH-1 se indicó en la **Figura I.4** de la Introducción.

GENERACIÓN DE SECUENCIAS CONSENSO EN 11 CS PARA CADA VARIANTE DEL VIH-1

Las secuencias consenso de cada variante son las que llevan el residuo más frecuente (nt o aa) en cada posición, tras hacer un alineamiento de múltiples secuencias en una variante viral concreta. Como un primer paso para este estudio se infirió la secuencia consenso (secuencia consenso inferida) en aa en la región codificante de *gag-pol* de cada variante del VIH-1, es decir, para cada uno de los 4 grupos del VIH-1, de los 9 subtipos y 51 CRF del grupo M del VIH-1, con secuencia disponible en GenBank en enero de 2012. Tras alinear todas las secuencias *gag/pol* nos centramos en el análisis de los 11 CS. Un segundo paso fue descargar y recuperar la secuencia consenso de GenBank para aquellas variantes en las que estuviera disponible (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>) como se explicó en el apartado Material y Métodos de esta Tesis. Por último, descargamos la secuencia

consenso de consensos proporcionada por GenBank con el fin de tomarla como referencia en un alineamiento realizado con las secuencias consenso individuales de *gag* y *pol*. Para generar esta secuencia consenso de consenso de *gag/pol* (incluyendo los 11 CS de interés), el GenBank alineó 28 secuencias consenso de *gag* (de 8 subtipos y 11 CRF) y 24 secuencias de *pol* (de 7 subtipos y 10 CRF) entre los 9 subtipos y 58 CRF descritos a fecha del estudio. La secuencia consenso de consensos fue tomada como referencia para el análisis de la conservación a nivel de aa entre variantes en los 110 residuos (10 aa en cada uno de los 11 CS) y fue utilizada para analizar el rango de conservación en los diferentes CS y variantes del VIH-1. La **Figura I.4** de la Introducción indica los 10 aa implicados en cada uno de los 11 CS y la secuencia consenso de consenso suministrada por GenBank para cada CS.

DETERMINACIÓN DEL GRADO DE CONSERVACIÓN DE CADA CS EN CADA VARIANTE DEL VIH-1

Tras un extenso análisis, comparamos nuestras secuencias consenso inferidas para cada variante del VIH-1 con la secuencia consenso de consensos de GenBank y conseguimos definir el grado de conservación en los 11 CS a nivel de aa. Determinamos los aa que diferían entre ambas secuencias consensos (inferida por nosotros vs. la recuperada de GenBank) para cada variante del VIH-1 con secuencias disponibles. El porcentaje exacto de residuos aminoacídicos conservados se calculó contando el nº de residuos coincidentes en cada una de las 10 posiciones de cada CS, en todas las secuencias atribuidas a una variante VIH-1 dada, dividido por el nº total de secuencias recuperadas para cada variante y multiplicada por los 10 residuos del CS, expresando el resultado en porcentaje. Para el análisis, establecimos un código de colores para representar el porcentaje de conservación a nivel de aa en cada sitio CS y variante VIH-1: verde ($\geq 90\%$), naranja (51-89%) y rojo ($\leq 50\%$).

ANÁLISIS ESTADÍSTICO

Los análisis estadísticos se realizaron utilizando el programa Epi Info v6.0 (Center for Disease Control and Prevention, Atlanta, GA, EE.UU.). Los cambios en las tasas fueron evaluados utilizando el análisis de Chi-cuadrado. El criterio para la significación estadística fue de $p < 0,05$.

RESULTADOS

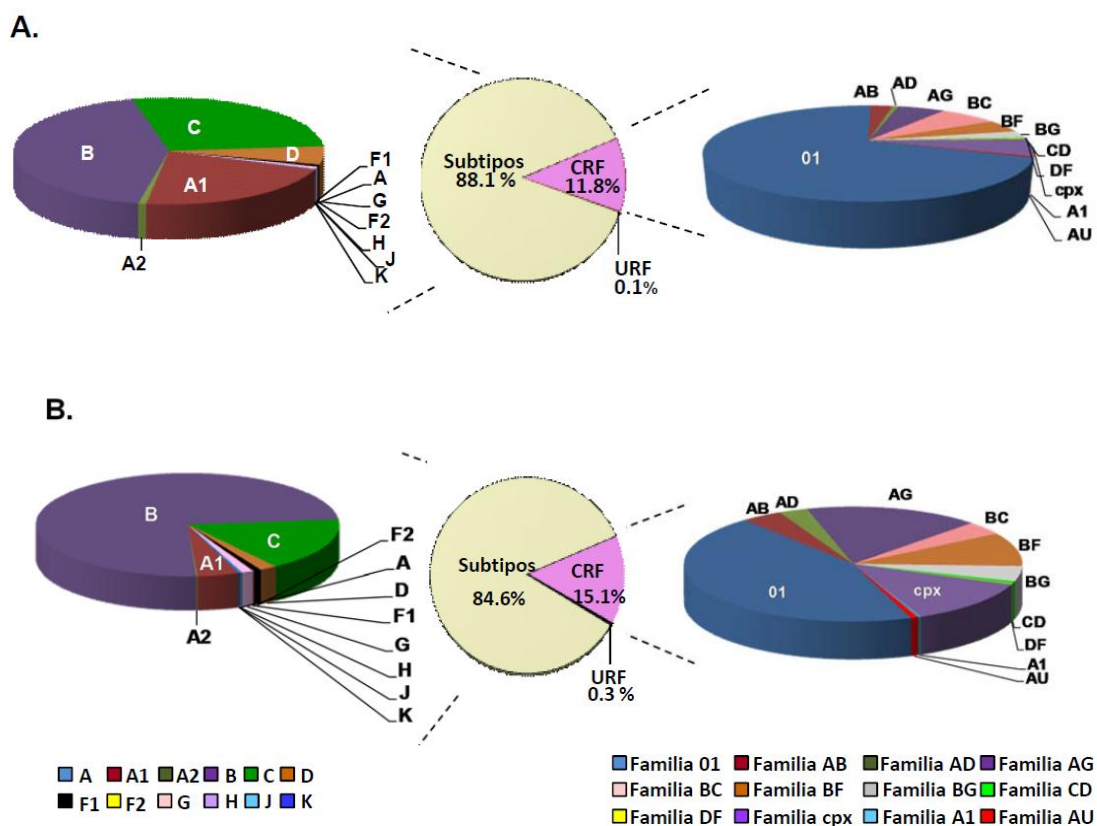
SECUENCIAS *gag/pol* VIH-1 EMPLEADAS EN EL ESTUDIO

En enero 2012 se descargaron 12.934 secuencias *gag/pol* (9.028 secuencias *gag* y 3.906 secuencias *pol* del VIH-1) de la base de datos de GenBank. Ya que cada secuencia tenía 2.844 nt, en el estudio analizamos 36.784.296 nt en total. Estos estaban localizados entre las posiciones 790-2.292 en *gag* y desde 2.253-5.096 en *pol* del aislado de referencia de subtipo B HXB2 (nº de acceso GenBank K03455), codificando las proteínas se muestran en la **Tabla I.1** de la Introducción.

La **Figura 1.1** muestra la distribución de variantes en el grupo M, de nuestras secuencias *gag/pol* de estudio recuperadas de GenBank. Estas secuencias pertenecían a los 4 grupos (M, O, N, P). En más detalle, 43/43 secuencias *gag/pol* pertenecían a los grupos O, N y P del VIH-1. Entre las 12.848 secuencias restantes del grupo M o mayoritario en la pandemia, un total de 8.985 pertenecían a *gag* (**Figura 1.1.A**) y 3.863 a *pol* (**Figura 1.1.B**), incluyendo representantes de todos los 9 subtipos y formas recombinantes (51 CRF y URF). Separando subtipos y los recombinantes CRF y URF dentro del grupo M del VIH-1, 7.913/3.269 (41,3%) secuencias *gag/pol* estaban adscritas a los 9 subtipos del grupo M del VIH-1 (A: sub-subtipos A1 y A2, B, C, D, F: sub-subtipos F1 y F2, G, H, J y K), y 1060/583 (55%) secuencias *gag/pol* a 51 CRF y 12/11 secuencias *gag/pol* a URF.

A pesar de la gran diferencia en el nº de secuencias recuperadas (8.985 *gag* versus 3.863 *pol*), la distribución específica de dichas secuencias en los subtipos del grupo M del VIH-1 y familias de CRF fue similar en ambos genes *gag/pol*. Entre los subtipos, las secuencias del subtipo B fueron las más representadas en ambas regiones codificantes *gag/pol* (43%/74,3%), seguido de secuencias pertenecientes al subtipo C (27,4%/16,7%), sub-subtipo A1 (21,6%/4,3%) y el subtipo D (5,6%/1,9%).

Figura 1.1. Distribución de variantes VIH-1 del grupo M entre las 9.028 secuencias *gag* (A) y 3.906 *pol* (B)



DISTRIBUCIÓN DE LOS CRF EN FAMILIAS DE RECOMBINANTES

Para facilitar el análisis posterior, los recombinantes CRF se agruparon en 12 familias de CRF estrechamente relacionados que compartían los mismos subtipos parentales y presentaban patrones de recombinación muy similares. La **Figura 1.2** muestra el patrón genómico de cada CRF mostrando los subtipos implicados su genoma viral, el nº de secuencias *gag* y *pol* que incluimos en el estudio de cada uno de ellos y también aquellas CRF con disponibilidad de secuencias consenso en GenBank para *gag* y *pol*. Dentro de los recombinantes CRF, la familia 01 fue la más representada tanto para *gag* (69,7%) como para *pol* (44,4%), seguida de las familias BC (7,2%), AG (5,9%), cpx (5,6%) y BF (4,5%) en *gag* y las familias AG (16,8%), cpx (12,5%), BF (9,4%), y BG (4,3%) en *pol*, entre otras. Las secuencias URF representaron menos del 0,3% del total de secuencias descargadas (12 secuencias *gag* y 11 secuencias *pol*).

Figura 1.2. Familias de recombinantes CRF del VIH-1 establecidas para este estudio

Variantes recombinantes VIH-1		Secuencias GenBank (no.)		Patrón mosaico subtipo									Secuencias consenso disponibles en GenBank		
Familia CRF	CRF	Gag	Pol	A	B	C	D	E	F	G	H	J	K	Gag	Pol
O1	CRF01_AE	737	256											SÍ	SÍ
	CRF22_01A1	2	3											NO	NO
A1	CRF32_06A1	2	2											NO	NO
AB	CRF03_AB	3	3											SÍ	SÍ
	CRF15_01B	4	5											NO	NO
	CRF33_01B	17	7											NO	NO
	CRF34_01B	0	3											NO	NO
	CRF48_01B	6	3											NO	NO
	CRF51_01B	0	2											NO	NO
AD	CRF16_A2D	1	2											NO	NO
	CRF21_A2D	0	3											NO	NO
	CRF35_AD	7	11											NO	NO
AG	CRF02_AG	59	94											SÍ	SÍ
	CRF43_02G	4	4											NO	NO
AU	CRF26_AU	4	4											NO	NO
BC	CRF07_BC	33	13											SÍ	NO
	CRF08_BC	40	6											SÍ	SÍ
	CRF31_BC	3	3											NO	NO
BF	CRF12_BF	7	5											SÍ	SÍ
	CRF17_BF	0	7											NO	NO
	CRF28_BF	3	3											NO	NO
	CRF29_BF	3	4											NO	NO
	CRF39_BF	3	3											NO	NO
	CRF40_BF	4	4											NO	NO
	CRF42_BF	16	17	No figure										NO	NO
	CRF44_BF	2	2											NO	NO
	CRF46_BF	8	8											NO	NO
	CRF47_BF	2	2											NO	NO

Variantes recombinantes VIH-1		Secuencias GenBank (no.)		Patrón mosaico subtipo									Secuencias consenso disponibles en GenBank		
Familia CRF	CRF	Gag	Pol	A	B	C	D	E	F	G	H	J	K	Gag	Pol
BG	CRF14_BG	11	11											SÍ	SÍ
	CRF20_BG	3	3											NO	NO
	CRF23_BG	2	2											NO	NO
	CRF24_BG	9	9											NO	NO
CD	CRF10_CD	3	3											SÍ	SÍ
cpx	CRF04_cpx	10	10											SÍ	SÍ
	CRF06_cpx	12	11											SÍ	SÍ
	CRF09_cpx	4	5	No figure										NO	NO
	CRF11_cpx	8	12											SÍ	SÍ
	CRF13_cpx	5	6											NO	NO
	CRF18_cpx	3	5											NO	NO
	CRF19_cpx	3	3											NO	NO
	CRF25_cpx	3	5											NO	NO
	CRF27_cpx	2	3											NO	NO
	CRF36_cpx	0	2											NO	NO
	CRF37_cpx	1	3											NO	NO
	CRF45_cpx	5	5											NO	NO
	CRF49_cpx	3	3											NO	NO
DF	CRF05_DF	3	3											NO	NO

nº, número; CRF, formas recombinantes circulantes; URF, formas recombinantes únicas. El **Anexo I** muestra una clasificación actualizada, incluyendo los 81 CRF descritos a la fecha de escritura de la Tesis (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>).

SECUENCIAS CONSENSO *gag/pol* DE CADA VARIANTE DEL VIH-1

Como ya se indicó en Material y Métodos, GenBank no tenía disponible la secuencia consenso para todas las variantes del grupo M del VIH-1, ni para los grupos O, N y P. La **Figura 1.2** muestra los recombinantes CRF del grupo M del VIH-1 con secuencia consenso disponible en GenBank en las regiones codificantes de *gag* y *pol*. GenBank no disponía de secuencia consenso en los CS para el sub-subtipo F2, ni para 40 de los 51 CRF con secuencia disponible a fecha de estudio (**Figura 1.2**). Tras agrupar todas las secuencias *gag* y *pol* de GenBank por variante de VIH-1, generamos nuestra propia secuencia consenso para los 4 grupos VIH-1 y todos los subtipos, sub-subtipos y recombinantes CRF del grupo M representados en las 9.028 secuencias Gag y 3.906 secuencias Pol del VIH-1 disponibles y descargadas de GenBank en enero de 2012, localizando los 110 residuos de los 11 CS. Una vez inferidas nuestras secuencias consenso por variante en cada CS, generamos las secuencias consenso de los 11 CS para todas las variantes del VIH-1, comparándolas con las secuencia consenso de cada variante dada por GenBank en caso de existir y generando nueva información en las variantes donde GenBank no disponía de dicha secuencia consenso.

La **Figura 1.3** indica las variantes del VIH-1 que mostraron cambios de aa respecto a la secuencia consenso de consensos de GenBank cuando aparecieron en una posición específica del CS en el 50% o más de las secuencias que descargamos de GenBank para esa variante concreta. Los asteriscos indican las variantes del VIH-1 con secuencia consenso no disponible en GenBank. En el cuadrado negro se indican los residuos de aa altamente conservados y presentes en más del 99% de las 9.028 secuencias Gag y 3.906 secuencias GagPol del VIH-1 con respecto a la secuencia consenso de consensos. Cuando dos residuos dentro de las secuencias analizadas de cada variante del VIH-1 mostraron una conservación del 50% las dos letras que hacen referencia a los residuos aminoacídicos localizados con la misma proporción aparecen en mayúsculas. Cuando tres o más residuos aparecían en la misma posición y ninguno de ellos presentaba una conservación de más del 50%, se mostraron en letras minúsculas y en orden de mayor a menor conservación.

Figura 1.3. Secuencias consenso en cada CS en variantes del VIH-1 con discrepancias respecto a la secuencia consenso de consensos de GenBank

Variantes VIH-1 en CS1		P17					P24				
		V	S	Q	N	Y	P	I	V	Q	N
Grupo M	02_AG, 32_06A1, 43_02G	S
	06_cpx	L
	14_BG	A
	27_cpx	VA	QH	.
	45_cpx, 49_cpx	I
Grupo N*		I	.	R	.	.	.	L	.	.	.
Grupo O*		T	g#	V	.	#1	.
#1, aosot.											

Variantes VIH-1 en CS2		P24					P2				
		K	A	R	V	L	A	E	A	M	S
Grupo M	F2*, URF*	.	.	.	I
	35_AD*	G
Grupo P* & O*		A

Variantes VIH-1 en CS3		P2					P7				
		T	T	—	—	I	M	M	Q	R	G
Grupo M	A*, A1, I, 35_AD*	.	N	—	—
	A2	.	N	—	#1
	B	S	A	—	T
	C	#2	N	—	S	.
	D	#3	A	—	A
	F1	.	#4	—	—	V	.	.	.	K	S
	F2*	TA	A	—	—	K	S
	G, 04_cpx, 13_cpx*	A	A	A	A	K	S
	H	A	N	AT	A	K	.
	K	PS	AV	—	—	V	.	.	.	RK	.
	01_AE	A	N	—	—
	02_AG	S	N	—	—
	03_AB, 32_06A1*	A	N	—	—	K	S
	05_DF*	S	#5	—	A	V
	06_cpx	#6	A	A	A	K	S
	07_BC	A	—	—	—	L	.	.	.	S	.
	08_BC	.	—	—	—	L	.	.	.	S	.
	09_cpx*	.	N	—	—
	10_CD	G	N	—	A
	11_cpx, 31_BC*	.	N	—	—	S	.
	12_BF, 46_BF*	.	.	—	—	V	.	.	.	K	S
	14_BG	A	.	—	—	K	S
	15_01B*	A	N	—	—	IV	.	.	.	GD	.
	16_A2D*	.	N	S	N	K	.
	18_cpx*	#7	N	—	—
	19_cpx*	S	#8	A	A	K	.
	20_BG*, 24_BG*	A	G	A	S	K	S
	22_01A1*, 33_01B*, 48_01B*	A	G	—	—
	23_BG*	A	G	A	S	.	ML	.	.	RK	S
	26_AU*	.	N	—	—	.	.	MV	.	.	.
	27_cpx*	A	AV	AT	A	K	SG
	28_BF*	A	A	—	—	V	.	.	.	K	S
	29_BF*	.	#9	—	—	V	.	.	.	K	S
	37_cpx*	P	.	—	—	.	.	.	K	Q	S
	39_BF*	Q	.	—	—	N
	40_BF*	.	A	—	—	V	.	.	.	K	S
	42_BF*	.	A	—	—	V	L	.	.	.	S
	43_02G*, 25_cpx*	A	A	A	A	K	.
	44_BF*	.	na	—	—	V	.	.	.	RK	S
	45_cpx*	T	N	—	—
	47_BF*	S	N	—	—	V	.	.	.	K	.
	49_cpx*	#10	#11	—	—
	URF*	.	#12	—	—
Grupo P*		S	K	.	.	R	R	V	YK	S	Q
Grupo N*		.	.	—	S	V	F	A	.	.	.
Grupo O*		#13	.	—	A	V	F	.	.	.	Q

Variantes VIH-1 en CS4		P7					P1				
		E	R	Q	A	N	F	L	G	K	I
Grupo M	03_AB	R	.
	05_DF*	V
	14_BG	.	S	K
	20_BG*, 23_BG*, 48_01B*	L
	37_cpx*	R	V
Grupo P*		G
Grupo N*		G	G
Grupo O*		G	K	Y

Variantes VIH-1 en CS5		P1					P6 ^{sig}				
		R	P	G	N	F	L	Q	S	R	P
Grupo M	A*, A1, A2	P
	F1, F2*, 06_cpx, 25_cpx*, 49_cpx*, 12_BF, 28_BF*, 29_BF*, 42_BF*, 46_BF*, 47_BF*, 14_BG, 32_06A1*	N	.	.
	G	N	.	#1
	01_AE, 02_AG, 22_01A1*, 09_cpx*, 45_cpx*, 37_cpx*, 03_AB, 35_AD*	P
	15_01B*	P	.	.	.	PL
	16_A2D*	P	.	.	.	T
	18_cpx*	K	.	.
	23_BG*, 26_AU*	P	.	N	.	LI
	24_BG*	P	.	N	.	L
	40_BF*	N	.	PL
	44_BF*	LP	.	N	.	PL
	48_01B*	I
Grupo P*		Y	V	.	K	Q	V
Grupo N*		P	.	T	T	K
Grupo O*		Y	#2	.	K	Q	V

#1, tnla; #2, tainvgkmpqler; #3, aglvsetpnm; #4, tsani; #5, pat; #6, tang; #7, tra; #8, nhs; #9, nta; #10, ggs; #11, shn; #12, antv; #13, elk.

Variantes VIH-1 en CS6		P7					TFP				
		E	R	Q	A	N	F	F	R	E	N
Grupo M	J*	D
	05_DF*	S
	08_BC	I
	14_BG	.	S	K	D
	15_01B*	#1
	20_BG*, 23_BG*, 24_BG*	T
	33_01B*	A
	40_BF*	ND
	48_01B*	K #2
		
Grupo P*		G	EK
Grupo N*		G	#3 G
Grupo O*		G	K	EQ I

#1, ntd; #2, adt; #3, ekq.

Variantes VIH-1 en CS8		P6 ^{pol}					PR				
		V	—	S	L	S	F	P	Q	I	T
Grupo M	A*	A	#1	.	F	.	FL
	A1	#2	#3	T	F
	A2	.	H	.	C	N
	B, D, 49_cpx*, 35_AD*	.	—	.	F
	C	G	—	T	.	N
	F1	.	P
	F2	VG	S	.	D
	G	#4	—	.	.	L
	H	—	—	.	N
	J*	.	—	.	S	N
Grupo M	K*	EG	SP	.	F	N
	01_AE	S	S	.	F
	02_AG	I	S	.	F	N
	03_AB	A	S	.	F	N
	04_cpx, 11_cpx, 07_BC*	I	—	.	F	N
	05_DF*, 12_BF, 17_BF*, 29_BF*	.	P
	06_cpx, 32_06A1*	I	—
	08_BC, 31_BC*	—	—	T	.	N
	09_cpx*	#5	P	.	F
	13_cpx*	I	—	.	F	N	C
Grupo M	14_BG	I	—	P	.	L
	15_01B*	#6	S	.	F
	16_A2D*	—	—	.	C	N
	18_cpx*	I	—	.	F	L
	19_cpx*	—	—	.	N	L
	20_BG*, 23_BG*, 24_BG*	—	—	G	.	N	L
	22_01A1*	A	S	.	F	.	L
	25_cpx*	I	—	.	#7	#8
	26_AU*	VA	—	.	F5
	27_cpx*	—	—	.	F
Grupo M	28_BF*	.	sp	.	.	#9
	33_01B*, 34_01B*, 48_01B*	I	—	.	F
	36_cpx*	I	S	.	.	N
	37_cpx*	I	P	.	F
	39_BF*, 21_A2D*	.	—	.	.	N
	40_BF*	.	P	.	.	SN
	42_BF*	.	P	.	.	N
	43_02G*	T	—	.	I	.	L
	44_BF*	.	P	.	.	SC
	45_cpx*	#9	P
Grupo M	46_BF*	#10	P
	47_BF*	VG	TS
	URF*	G	S	.	F	N
		GV	IV	P	F	SN	L
	
	
	
	
	
	
	
Grupo P*		.	P	T	.	N
Grupo N*	
Grupo O*		L	—	.	V	C	L	.	.	.	P

#1, lgi; #2, gaetdvslm; #3, pslyfghirt; #4, ltn; #5, atv; #6, yes; #7, ycfi; #8, gdc; 9, sng; #10, lqiv

Variantes VIH-1 en CS7		TFP					P6 ^{pol}				
		E	N	L	A	F	Q	Q	G	E	A
Grupo M	A*, H, 09_cpx*, 13_cpx*, 25_cpx*, 45_cpx*, 03_AB	R	.	.
	B, D	P	.	.	K	.
	C, 29_BF*, 43_02G*, 31_BC*, 28_BF*	P
	J*	.	D	R	.	.
	01_AE	K	.
	04_cpx	.	.	V	R	.	.
	05_DF*	.	S	.	.	.	P
	07_BC*	L	P	.	.	K	.
	08_BC	.	I	.	.	.	P
	10_CD	R	K	.
Grupo M	15_01B*	.	#1	.	.	.	QP	.	.	K	.
	20_BG*, 24_BG*	.	T	.	.	.	P	.	.	K	.
	23_BG*	.	T	.	.	.	P	.	.	EK	.
	33_01B*	.	A	.	.	.	P
	34_01B*	.	I	K	.
	40_BF*	.	ND	.	.	FS	PL
	42_BF*, 21_A2D*	#2
	48_01B*	K	#3	.	.	.	P
	51_01B*	.	D	.	.	.	P	.	.	K	.
		EK	D	.	.	S	WG	G	Q	.	.
Grupo P*		#4	G	.	V	S	L	.	R	.	T
Grupo N*		EQ	I	.	.	S	G	G	H	.	.
Grupo O*											

#1, ntd; #2, lqj; #3, adt; #4, ekq.

Variantes VIH-1 en CS9		PR					RT ^{pol}				
		C	T	L	N	F	P	I	S	P	I
Grupo M	11_cpx	IV
	16_A2D	IV	.	PT	.
	32_06A1	FL
Grupo P*		.	.	.	S	PS	.

Variantes VIH-1 en CS10		RT ^{pol}					RT ^{pol}				
		G	A	E	T	F	Y	V	D	G	A
Grupo M	G, H, 14_BG, 13_cpx	Y
	07_BC*, 08_BC, 32_06A1*
	17_BF*	E
	26_AU*	FY
	36_cpx*	.	.	EG	TR
	43_02G*	.	AV
		.	IT
		Y
	
	
Grupo P*		.	IT
Grupo N* & O*		Y

Variantes VIH-1 en CS11		RT ^{pol}					IN				
		I	R	K	V	L	F	L	D	G	I
Grupo M	F1, 36_cpx*, 05_DF*, 17_BF*, 28_BF*, 29_BF*, 42_BF*, 46_BF*, 12_BF
	03_AB	.	.	#1
	23_BG*	.	.	KR
	51_01B*	.	.	R
		.	.	R	E	.	.
	
	
	
	
	
Grupo P* & O*		.	.	R	E	.	.

#1, kre.

Numeración en aa de los CS considerando como aa 1 el primer residuo de la poliproteína Gag Pr55^{Gag}: CS1, 128-137; CS2, 359-368; CS3, 373-382; CS4, 428-437; CS5, 444-453; CS6, 427-TFP; CS7, TFP-P6^{pol}; CS8, P6^{pol}-151; CS9, 151-160; CS10, 591-600; CS11, 711-720.

Así también identificamos las variantes del VIH-1 que presentaban secuencias consenso en algunos CS diferentes a la inferida por GenBank (**Figura 1.3**), proporcionando nueva información no existente hasta este estudio con posible relevancia biológica o clínica.

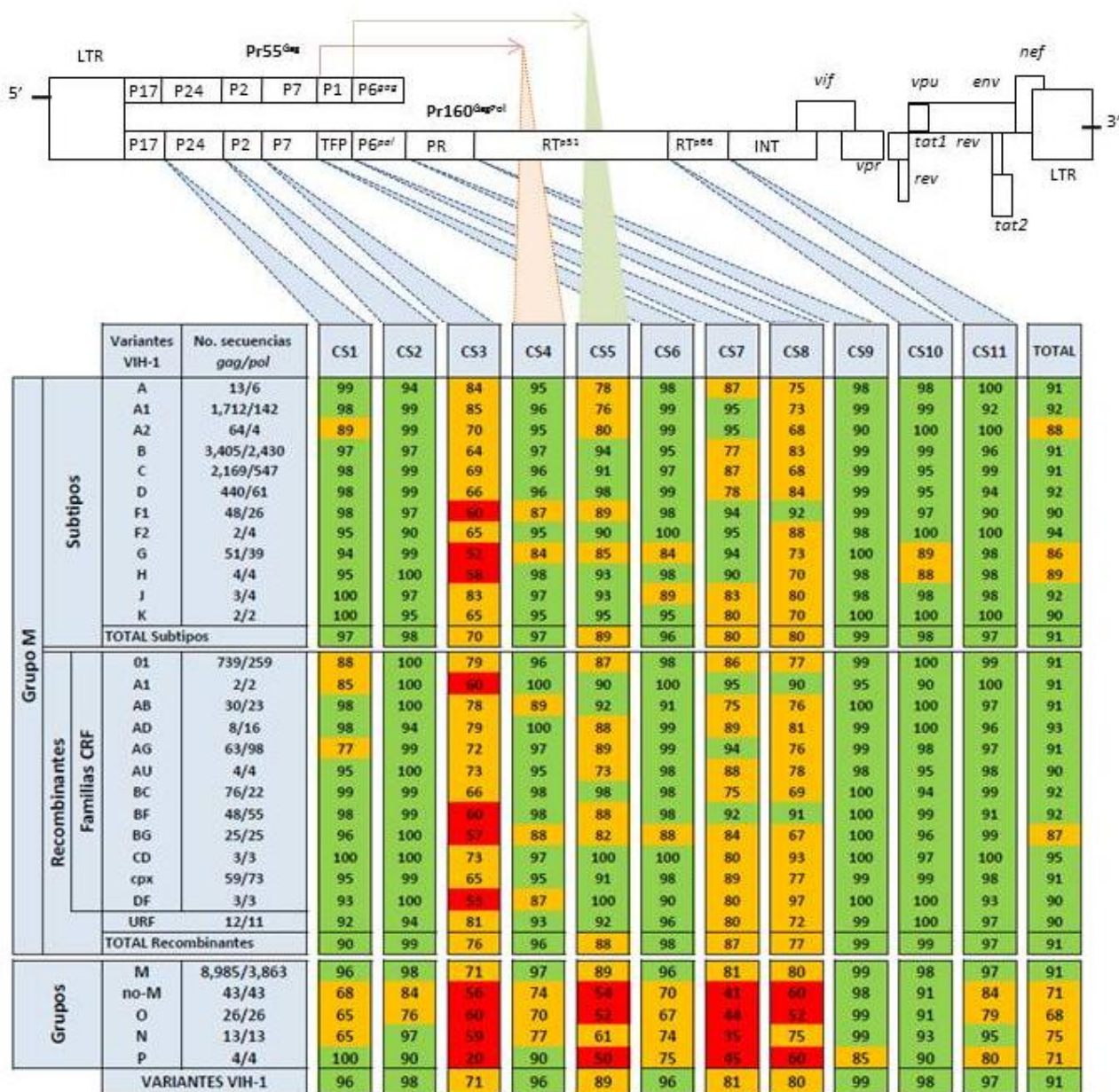
IDENTIFICACIÓN DE LOS RESIDUOS ALTAMENTE CONSERVADOS EN LOS 11 CS ENTRE VARIANTES DEL VIH-1

Tras localizar los 11 CS en los alineamientos de aa, comparamos el grado de conservación, a nivel de aa, en cada CS, determinado por el nº de aa coincidentes entre los 10 residuos de cada CS de nuestras secuencias consenso inferidas para cada variante del VIH-1 a partir de las 12.934 secuencias descargadas con respecto a la secuencia consenso de consensos proporcionada por GenBank.

La **Figura 1.4** muestra el grado de conservación a nivel de aa de los 10 residuos de cada CS dentro de los precursores virales Gag (Pr55) y Gag-Pol (Pr160) en todas las variantes VIH-1 analizadas, incluyendo secuencias de los 4 grupos, 9 subtipos y 51 CRF del grupo M. Evaluamos el porcentaje de residuos conservados en cada variante VIH-1 y en cada CS como explicamos en los Métodos de este capítulo. Establecimos un código de colores para indicar diferentes niveles de conservación a nivel de aa en cada sitio CS y en todas las secuencias correspondientes a cada variante VIH-1: verde ($\geq 90\%$), naranja (51-89%) y rojo ($\leq 50\%$).

Por tanto, en verde estarían aquellos aa que se conservan al menos más de un 90% en el global de todas las secuencias descargadas. Éstas serían posiciones de extremo interés en investigación para el diseño de cebadores y sondas para la detección molecular del VIH-1 o como potencial diana farmacológica futura. También identificamos 52/110 (47.3%) aa conservados en más del 99% de las 9.028 secuencias *gag* y 3.906 secuencias *pol* del VIH-1 con respecto a la secuencia consenso de consensos, residuos que ya habíamos resaltado en negrita en la **Figura 1.3**. Por tanto, cerca de la mitad (52,7%) de los 110 aa dentro de los 11 CS podrían admitir un grado diferente de variabilidad variante dependiente.

Figura 1.4. Conservación de los 11 sitios de procesamiento viral o CS, a nivel de aa, localizados en los precursores Gag y GagPol, en las variantes VIH-1 analizadas



Niveles de conservación a nivel de aa en cada sitio CS y en todas las secuencias correspondientes a cada variante VIH-1: verde ($\geq 90\%$), naranja (51-89%) y rojo ($\leq 50\%$). CS, cleavage site; P17, matriz; P24, cápside; P2 espaciador peptídico 1; P7, nucleocápside; P1, espacio peptídico 2; TFP, proteína marco de lectura (*transframe protein*); PR, proteasa; RT, retrotranscriptasa; IN, integrasa; CRF, forma recombinante circulante; URF, forma recombinante única.

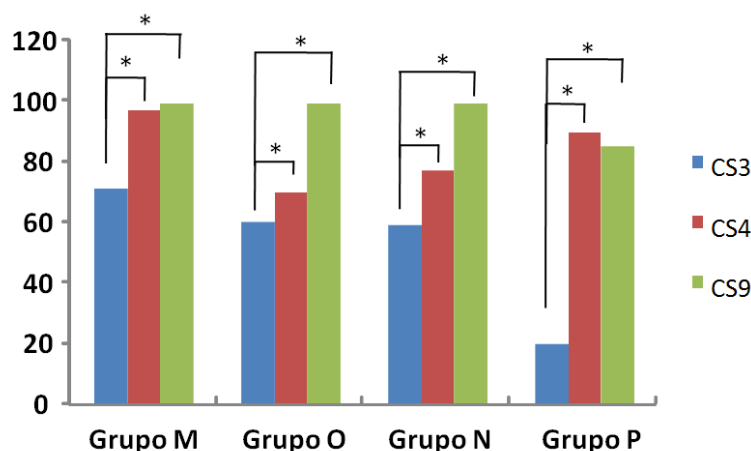
DIFERENCIAS OBSERVADAS EN EL GRADO DE CONSERVACIÓN DE LOS 11 CS ENTRE VARIANTES DEL VIH-1

Entre sitios de escisión proteolítica

A pesar del papel estructural y funcional de las proteínas en el ciclo viral, observamos diferentes grados de conservación entre los 11 CS y las distintas variantes del VIH-1. El CS con el nº más alto de residuos conservados fue el CS9 (PR/RTp51, 99%), seguido de CS10 (RTp51/RTp66, 98%), CS2 (P24/P2, 98%), CS11 (RTp66/IN, 97%), CS1 (P17/P24, 96%), CS4 (P7/P1, 96%) y CS6 (P7/TFP, 96%). Los CS menos conservados entre los grupos del VIH-1 y subtipos y recombinantes del grupo M fueron CS3 (P2/P7, 71%), CS8 (P6pol/PR, 80%), CS7 (TFP/P6pol, 81%) y CS5 (P1/P6gag, 89%).

Considerando el orden de procesamiento proteico previamente definido (de Oliveira *et al.*, 2003; Pettit *et al.*, 2005; Fun *et al.*, 2012), al no cortarse todos los CS en el mismo orden por la PR viral, observamos curiosamente una conservación mayor de manera significativa en los últimos sitios de procesamiento en los precursores Gag (CS4, P7/P1) y GagPol (CS9, PR/RTp51) comparados con el primer sitio de procesamiento descrito (CS3, P2/P7) en los 4 grupos del VIH-1 (**Figura 1.5**). De este modo comprobamos como el nivel de conservación encontrado es distinto conforme a las etapas de procesamiento.

Figura 1.5. Conservación de los primeros y últimos sitios de procesamiento entre grupos del VIH-1



CS3 (PR/RTp51), primer sitio de procesamiento; CS4 (P7/P1) y CS9 (PR/RTp51), últimos sitios de procesamiento por la PR viral. Asterisco, diferencia significativa, con $p < 0,01$.

Entre grupos del VIH-1

También observamos diferencias en el grado de conservación en los CS entre grupos VIH-1 (**Figura 1.4**). El CS10 (RT^{p51}/RT^{p66}) mostró más del 90% de los residuos conservados respecto a la secuencia aminoacídica de consenso de consensos, en los 4 grupos VIH-1. Comparando los grupos no-M y M, observamos una elevada y similar conservación en CS9 (99% y 98%, respectivamente) y en CS10 (98% y 91%, respectivamente). Sin embargo, CS7 (TFP/P6^{pol}) presentó el menor rango de conservación entre los grupos no-M (41%), seguido de CS5 (P1/P6^{gag}, 54%), CS3 (P2/P7, 56%), CS8 (P6^{pol}/PR, 60%) y CS1 (P17/P24, 68%). El grupo O mostró la menor conservación en 6 de los 11 CS (CS1, CS2, CS4, CS6, CS8 y CS11), el grupo N en CS1 y CS7 y el grupo P en CS3, CS5, CS9 y CS10 (**Figuras 1.3 y 1.4**). Considerando los 11 CS, encontramos un grado de conservación significativamente mayor entre el grupo M vs. las distintas variantes correspondientes a los otros grupos, grupo no-M (91% vs. 71%, $p < 0,001$), probablemente debido al uso del grupo M como consenso para comparar.

Entre subtipos y recombinantes del Grupo M del VIH-1

Siete de los once sitios de procesamiento (CS1, CS2, CS4, CS6, CS9, CS10 y CS11) estaban bien conservados dentro del total de subtipos y recombinantes del grupo M del VIH-1, mostrando más del 90% de conservación (**Figura 1.4**). Sin embargo, los cuatro CS restantes (CS3, CS5, CS7 y CS8) mostraron un grado de conservación más variable en un gran nº de variantes VIH-1. El grado más bajo de conservación en los 11 CS fue encontrado en los siguientes subtipos y recombinantes: CS1 (P17/P24) en sub-subtipo A2 (89%) y en la familia de recombinantes AG (77%); CS2 (P24/P2) en sub-subtipo F2 (90%) y familia de recombinantes AD y URF (ambos 94%); CS3 (P2/P7) en subtipo G (52%) y familia recombinante DF (87%); CS5 (P1/P6^{gag}) en sub-subtipo A1 (76%) y familia recombinante AU (73%); CS6 (P7/TFP) en subtipo G (84%) y familia recombinante BG (88%); CS7 (TFP/P6^{pol}) en subtipo B (77%) y familias recombinantes AB y BC (ambos 75%); CS8 (P6^{pol}/PR) en sub-subtipo A2 y subtipo C (ambos 68%) y familia recombinante BG (67%); CS9 (PR/RT^{p51}) sub-subtipo A2 (90%) y familia recombinantes A1 (95%); CS10 (RT^{p51}/RT^{p66}) en subtipos H (88%) y G (89%) y familia recombinante A1 (90%); y CS11 (RTP^{p66}/IN) en sub-subtipo F1 (90%) y familia recombinante BF (91%). De este modo, el

subtipo G mostró la más alta variabilidad en CS3, CS4 y CS6 y el subtipo B en CS7 comparado con otros subtipos del grupo M.

Las familias de recombinantes DF (53%), BG (57%), A1 (60%) y BF (60%) mostraron la más alta variabilidad en CS3. Los CS2, CS6, CS9, CS10 y CS11 fueron altamente conservados (94-99%) entre todas las familias CRF y URF.

En general, los recombinantes mostraron la misma conservación que los subtipos (91%) en los 11 CS analizados, la familia de recombinantes BG con 87%, subtipo G con 86%, sub-subtipo A2 con 88% y subtipo H con 89%. Sin embargo, observamos que el grado de conservación fue significativamente más bajo en los subtipos que en los recombinantes en CS3 (70% vs. 76%, $p<0,001$) y en CS7 (80% vs. 87%, $p<0,001$) y mayor en CS1 (97% vs. 90%, $p<0,001$) (**Figura 1.4**).

Entre sub-subtipos del grupo M del VIH-1

El sub-subtipo A2 presentó significativamente menor conservación que el sub-subtipo A1 en CS1 (89% vs. 98%, $p<0,001$), en CS3 (70% vs. 85%, $p<0,001$) y en CS9 (90% vs. 99%, $p<0,001$). Mientras que la conservación fue significativamente mayor conservación en CS5 al comparar ambos sub-subtipos (80% vs. 76%, $p=0,04$). La conservación fue de gran significancia en CS11 (100% vs. 92%, $p=0,06$).

El sub-subtipo F2 mostró una conservación superior comparado con el sub-subtipo F1 en CS11 (100% vs. 90%, $p=0,04$), aunque el nº de secuencias F2 disponibles fue muy bajo (ver **Figura 1.4**).

DISCUSIÓN

El análisis de genomas VIH-1 proporciona información biológica de gran utilidad para conocer mejor la estructura y la función de las proteínas virales (Doherty *et al.*, 2005). La maduración del virus es esencial para su infectividad, y ello depende directamente del procesamiento proteolítico correcto de las poliproteínas precursoras Gag y GagPol. Mutaciones que interrumpan o que alteren el orden de la escisión o corte en los sitios de escisión o CS, producen partículas virales aberrantes que poseen una infectividad significativamente reducida (Adamson, 2012).

A pesar de que otras publicaciones han expuesto de manera previa que determinados CS estaban más conservados que otros, en estos estudios sólo se analizó un nº limitado de variantes VIH-1 y de secuencias de CS (de Oliveira *et al.*, 2003; Luo *et al.*, 2013; Liégeois *et al.*, 2013). Nuestro estudio es el primero que evalúa el rango de conservación en 11 CS dentro de los precursores Gag y GagPol, y que define las secuencias consenso de cada CS. Ello fue posible porque utilizamos una gran base de datos de secuencias (GenBank) e incluimos todos los subtipos del grupo M y todos los CRF con secuencia disponible, cuya prevalencia y complejidad está viéndose incrementada en la pandemia (Zhuang *et al.*, 2002). Asimismo, es el primer estudio que incluye secuencias de los grupos O, N y P, identificando los residuos conservados completamente en los CS presentes en los 4 grupos del VIH-1. Nuestros resultados revelan que el rango de conservación encontrado en cada variante VIH-1 y en cada CS difiere, tanto entre los distintos sitios de escisión estudiados, como entre los 4 grupos VIH-1 y las variantes del grupo M, subtipos y recombinantes.

Observamos un nº distinto de secuencias *gag* y *pol* recuperadas de GenBank en enero de 2012 y empleadas en este trabajo. Ello podría ser explicado porque, a pesar de que el gen *pol* es el más estudiado en el VIH y secuenciado para detectar resistencias virales, la mayoría de las veces esos resultados sólo se usan para informar de resistencias durante el seguimiento clínico del paciente y las secuencias no se depositan en el GenBank. Ello explicaría el menor nº de secuencias *pol* disponibles respecto a las secuencias *gag*. Éstas últimas se generan fundamentalmente para trabajos de investigación por su importancia funcional y estructural, y normalmente se depositan en el GenBank antes de la publicación de resultados en revistas científicas, siendo obligatorios los números de accesos antes de publicarlas.

NUEVOS HALLAZGOS EN LA VARIABILIDAD DE LOS CS ENTRE VARIANTES

Hasta el momento únicamente un nº limitado de estudios habían evaluado la variabilidad natural entre variantes del VIH-1 dentro del gen *gag* y los CS (Bally *et al.*, 2000; de Oliveira *et al.*, 2003; Malet *et al.*, 2007; Snoeck *et al.*, 2011; Fun *et al.*, 2012; Luo *et al.*, 2013). Estos estudios se habían enfocado principalmente en el subtipo B y/o subtipo C analizando, en la mayoría de casos, un pequeño nº de secuencias y/o de CS.

Además, la mayoría de estos trabajos habían utilizado secuencias del subtipo B como secuencia de referencia para el análisis de la conservación, bien del aislado HXB2 (Malet *et al.*, 2007; Snoeck *et al.*, 2011; Fun *et al.*, 2012), o del aislado pNL-4-3 (Bally *et al.*, 2000), o incluso habían usado la secuencia ancestral común mas reciente del Grupo M (de Oliveira *et al.*, 2003). Sólo uno de estos estudios había empleado la secuencia consenso de consensos proporcionada por GenBank como referencia para establecer las comparaciones y evaluar la conservación de los distintos residuos (Luo *et al.*, 2013). Cabe destacar también que ninguno de estos estudios previos había inferido una secuencia consenso para cada variante VIH-1 y para cada región de procesamiento o CS. Si bien no hay que obviar que algunos de los otros trabajos, al igual que nosotros, habían incluido secuencias pertenecientes a algunas CRF (pero en bajo nº) o a algunas variantes distintas al grupo M.

A pesar de la llamativa diferencia entre el nº de secuencias que descargamos de GenBank para el grupo M (8.985 secuencias *gag*/3.863 secuencias *pol*), con respecto al resto de grupos (43/43 secuencias *gag/pol*) o determinados subtipos (H, J, K), sub-subtipos (F2) o CRF, los datos disponibles permiten establecer comparaciones entre los rangos de conservación en los distintos CS analizados. Destacar también que este trabajo nos permitió definir diferencias específicas para cada secuencia consenso cuando nos fijamos en la región de CS entre las distintas variantes VIH-1, grupo, subtipo, CRF y URF. Nuestros datos reflejan que el grado de conservación difiere entre posiciones de aa individuales en los CS y proporciona discrepancias significativas entre variantes específicas VIH-1 y entre sitios de procesamiento, mejorando así los datos disponibles en GenBank para secuencias consenso de variantes VIH-1 específicas.

Empleando un gran conjunto de datos (12.934 secuencias) de todas las variantes VIH-1, nuestro estudio reveló que los CS3, CS5, CS7 y CS8 fueron los sitios de procesamiento menos conservados entre todas las variantes VIH-1. Este hallazgo está de acuerdo con publicaciones previas que habían utilizado un nº de secuencias notablemente inferior, 32 secuencias del subtipo C, 34 del subtipo B y 18 secuencias de otros subtipos (de Oliveira *et al.*, 2003).

Somos conscientes, y así quedo reflejado en la publicación de este trabajo, que son necesarios estudios adicionales para entender la alta variabilidad en los CS, los cuales son esenciales en el ciclo viral. En concreto, CS3 es el primer sitio de procesamiento en los precursores Gag y GagPol y es crítico para la maduración del dímero ARN (Shehu-Xhilaga *et al.*, 2001). El CS7 está implicado en la activación de la PR viral y en el momento y especificidad de la escisión de GagPol (Tomasselli *et al.*, 1994). El CS5 es el responsable de la síntesis de la proteína P6^{gag}, la cual se requiere para la maduración y liberación del virión infeccioso (Müller *et al.*, 2002). El CS8 representa una pieza clave en el autoprocesamiento de la PR y además podría estar implicado en la adecuada dimerización que requiere la PR (Sadiq *et al.*, 2012).

LIMITACIONES ESTRUCTURALES DE LA VARIACIÓN EN LOS SITIOS DE ESCISIÓN O CS

Para una escisión eficiente de Gag y GagPol por parte de la PR, se requieren interacciones complejas de sustratos aminoacídicos localizados en los sitios activos de la PR viral. La PR del VIH-1 sólo es funcional en su forma dimérica y precisa que un solo monómero sea introducido dentro de cada precursor. Dos monómeros individuales, en diferentes cadenas GagPol, deben entonces permanecer juntos para formar una PR dimérica, que se unirá finalmente consigo misma para alcanzar la forma madura activa (Sadiq *et al.*, 2012). La maduración del VIH-1 requiere el reconocimiento por parte de la PR de la conformación tridimensional y asimétrica del sustrato Gag, en lugar de una secuencia peptídica concreta (Prabu-Jeyabalan *et al.*, 2002). Posteriormente, la PR media la escisión de las poliproteínas estructurales del VIH-1 Gag y GagPol, mediante la interacción con CS específicos (Kaplan *et al.*, 1994; Adamson *et al.*, 2012). Cada sustrato tiene una estructura única, que difiere en la composición de aa (de Oliveira *et al.*, 2003). Se piensa que esas pequeñas diferencias en la estructura del sustrato, afectan a la afinidad de la PR viral y contribuyen a un proceso de maduración gradual altamente regulado y ordenado, en el cuál, las escisiones individuales tienen lugar a diferentes tiempos y proporción (de Oliveira *et al.*, 2003; Pettit *et al.*, 2005). En el procesamiento de Gag y GagPol están implicados determinantes adicionales, además de secuencias aminoacídicas y estructuras secundarias locales de CS (Lee *et al.*, 2012). Existen limitaciones en la viabilidad de las cepas virales que presentan múltiples mutaciones, debido al hecho de que la existencia de mutaciones combinadas pueden desestabilizar las

interacciones estructurales de las multiproteínas que son críticas para la función viral (Dahirel *et al.*, 2011). Por tanto, la conservación de secuencias aminoacídicas, indicaría que estos aa específicos se requieren para mantener la estructura básica y función, aunque algunos autores han sugerido un importante papel de la estructura del ARN en la conservación del VIH-1 (Snoeck *et al.*, 2011; van der Kuyl *et al.*, 2012). Se sabe que las propiedades físicas, químicas y estructurales de determinadas proteínas del VIH-1 con papeles funcionales en el ciclo viral, como gp41, pueden estar altamente conservadas, todo ello a pesar de una notable diversidad presente a lo largo de las secuencias. Esto parecería por tanto indicar un delicado balance entre las presiones evolutivas y la conservación de la estructura de la proteína (Steckbeck *et al.*, 2011). Por último señalar que la estructura de la proteína, en concreto de los dominios α -hélice, han sido asociados con la conservación del VIH-1 (Snoeck *et al.*, 2011) y es un elemento estructural estable entre las proteínas (Richardson, 1981).

Además, la conservación significativamente mayor, observada comparando el último CS presente en *gag* vs. el primero, en los precursores Gag y GagPol, podría sugerir la existencia de un requisito más alto en el nº de limitaciones estructurales en los últimos pasos del procesamiento viral.

Aunque el objetivo de nuestro estudio es puramente descriptivo, creemos que podría utilizarse como herramienta de trabajo en investigación, para un mejor entendimiento de la estructura de los CS. Estos conocimientos son fundamentales para entender la importancia que tiene dicha estructura para que tenga lugar una escisión correcta en las distintas variantes VIH-1. Se necesitan más estudios para comprender si la presencia de residuos específicos dentro de los CS para cada variante VIH-1 modularían o no, la capacidad replicativa de la variante correspondiente, lo que previamente se había observado para polimorfismos naturales específicos en la PR en algunas variantes no-B (Holguín *et al.*, 2006b).

Se ha sugerido que secuencias localizadas alrededor de los CS en *gag* están igualmente conservadas que otros motivos funcionales. Mientras que secuencias que son diana de inhibidores de la RT están más conservadas que otros motivos no-funcionales (Luo *et al.*, 2013).

Todos estos resultados se han publicado bajo el título “New findings in cleavage sites variability across groups, subtypes and recombinants of human immunodeficiency virus type 1”. PLoS One. 2014; 9(2):e88099. Factor de impacto: 3,73. Cuartil: 1. <http://www.ncbi.nlm.nih.gov/pubmed/24516589>

CAPÍTULO 2

**Variabilidad natural en posiciones asociadas
a resistencia a inhibidores de la proteasa
(MRD-IP) en Gag y PR de distintas variantes
del VIH-1. Efecto del TAR en algunas MRD-IP
en Gag**

“En la vida hay algo peor que el fracaso: el no haber intentado nada”

Franklin D. Roosevelt (1822-1945)

JUSTIFICACIÓN Y OBJETIVOS

El genoma del VIH difiere entre las distintas variantes del VIH-1. Como hemos dicho anteriormente en la Introducción de esta Tesis, existen polimorfismos naturales a lo largo del genoma que son específicos de una o varias variantes del VIH-1 y fijados en el genoma durante la evolución viral. Por ello pueden servir como marcadores de grupos, subtipos y recombinantes CRF. Algunos de esos polimorfismos se localizan en posiciones de proteínas donde se ubican los cambios asociados a resistencias a fármacos antirretrovirales (ARV). La disponibilidad de secuencias genéticas en la base de datos del GenBank de pacientes que no han recibido ARV (pacientes *naïve*) permite estudiar la variabilidad natural (en ausencia de fármacos) en esas posiciones asociadas a resistencia a ARV.

Como se describió en el apartado de Introducción de esta Tesis, la resistencia a fármacos inhibidores de la proteasa (IP) en pacientes en fracaso virológico (FV) se ha asociado principalmente a mutaciones en la PR viral, pero también a cambios en los genes *gag*. Estos últimos pueden favorecer la resistencia a IP, incluso en ausencia de mutaciones en la PR, siendo propuesto como un nuevo mecanismo de resistencia para IP. Así, mutaciones en *gag* pueden compensar los defectos de capacidad replicativa de virus con una PR viral resistente a IP menos procesiva.

En la actualidad disponemos de pocos estudios genéticos que presten atención a la variabilidad natural de *gag* y de su relación con resistencia y fracasos terapéuticos a IP para su aplicación clínica. Por ello en este capítulo nos propusimos:

- Analizar los polimorfismos naturales a nivel de aa de la PR y Gag asociados a resistencia a IP (MRD-IP) en distintas variantes del VIH-1 recuperadas de pacientes *naïve* y con secuencia disponible en el GenBank.
- Identificar si el TAR con IP afecta a patrones mutacionales en algunas posiciones de proteínas Gag relacionados con MRD-IP empleando secuencias del VIH-1 recuperadas de muestras clínicas de pacientes con o sin exposición previa a IP.

MATERIAL Y MÉTODOS

SECUENCIAS GAG Y PR DE PACIENTES *NAIVE* DESCARGADAS DE GENBANK

Para llevar a cabo los objetivos del presente capítulo, en una primera fase desarrollada en junio de 2015 recuperamos del GenBank todas las secuencias GagPol disponibles para cada grupo (O, N, P y M) y subtipo o CRF del grupo M del VIH-1 que pertenecían a pacientes *naive* a ARV. El procedimiento de descarga de secuencias se indicó en el apartado de Materiales y Métodos, seleccionando las secuencias pertenecientes a pacientes *naive* para lo que se empleó el filtro “*only drug naive sequences*” previamente a la descarga de secuencias en el GenBank. De esta manera se descargaron un total de 2.141 secuencias GagPol de VIH-1 de 716 pacientes *naive*. El objetivo fue evaluar la variabilidad asociada a cada variante viral en posiciones de interés en las proteínas Gag y PR relacionadas con la resistencia a fármacos IP. El procedimiento para alinear las 2.141 secuencias con respecto a la secuencia “consenso de consensos” descargada del GenBank empleando el programa Mega v5.0 se explicó previamente en el apartado de Materiales y Métodos.

SECUENCIAS DEL VIH DE MUESTRAS CLÍNICAS EMPLEADAS PARA ANALIZAR EL EFECTO DEL TAR EN ALGUNAS POSICIONES DE GAG RELACIONADAS CON MRD-IP

Queríamos identificar si el TAR con IP afectaba a los patrones mutacionales en posiciones de Gag localizadas en 4 de las proteínas Gag (P2, P7, P1 y P6) previamente relacionados con MDR-IP. Para ello, recuperamos secuencias *gag* de virus de muestras clínicas pertenecientes a 149 pacientes infectados por diversas variantes del VIH-1 y en ausencia o presencia de TAR con IP, mayoritariamente en seguimiento clínico en el Hospital Carlos III de Madrid antes del año 2013. Dichas secuencias estaban disponibles en el Laboratorio de Epidemiología Molecular del VIH (EpiMolVIH), adscrito al Departamento de Microbiología y Parasitología del Hospital Universitario Ramón y Cajal y al Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS) y al CIBER de Epidemiología y Salud Pública (CIBERESP). La mayoría de ellas (122/149, 81,9%) habían sido previamente publicadas y depositadas en el GenBank (Holguín *et al.*, 2005). Las secuencias *gag*

analizadas de 515 pares de bases incluían la región C-terminal de P24, P2, P7, P1 y P6^{gag} (posiciones 1754–2269 del aislado HXB2; GenBank K03455).

Las 149 secuencias Gag disponibles recuperadas de pacientes con y sin experiencia previa a IP se alinearon a nivel de nt usando el programa Mega v5.0 con la secuencia “consenso de consensos” en *gag* descargada del GenBank (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>), tal y como se describió previamente en Métodos. El alineamiento se llevó a cabo agrupando las secuencias por variantes del VIH. Tras localizar las posiciones asociadas a MRD-IP en esas secuencias de 515 nt, se calculó el grado de conservación de cada MRD-IP en la población de estudio *naïve* y tratada con IP.

FUENTES DE LOCALIZACIÓN DE RESIDUOS EN PR Y GAG RELACIONADOS CON RESISTENCIA A IP

Una vez descargadas y alineadas todas las secuencias, procedimos a estudiar la presencia de los cambios a nivel de aa de las proteínas Gag y PR asociados a resistencias a IP (MRD-IP) y previamente descritos en diversas fuentes de información. En la PR se tomaron como referencia las MRD-IP descritas por el grupo IAS-USA en 2015 (Wensing *et al.*, 2017) y la base de datos de resistencias a fármacos VIH de la Universidad de Stanford (HIV Drug Resistance Database, <https://hivdb.stanford.edu/dr-summary/resistance-notes/PI/>).

Así, estudiamos la variabilidad natural en un total de 96 MRD-IP ubicadas en 37 residuos de la PR viral. De esos 96 cambios aminoacídicos presentes en la PR en nuestras secuencias analizadas, 36 de ellos se ubicaron en 12 posiciones asociadas a resistencia primaria a IP y 43 cambios se localizaron en 24 posiciones relacionadas con resistencia secundaria a IP. Los 17 cambios restantes estaban en 7 posiciones de resistencia no primarias ni secundarias descritas por Stanford como “otras mutaciones relacionadas con IP”. Algunas de las últimas son altamente polimórficas y asociadas pobremente a resistencia a IP, incrementando la capacidad replicativa o *fitness* del virus cuando aparecen en combinación con MRD primarias (<https://hivdb.stanford.edu/page/release-notes/#drug.resistance.mutations.drms.and.sequence.interpretation>).

Es importante recordar que la PR tiene seis posiciones (10, 20, 46, 74, 82, 88) que pueden admitir cambios asociados a más de un grupo de MRD-IP (primarias, secundarias o raras). Cada MRD-IP en la PR recibe una puntuación de penalización o “*penalty score*” respecto al nivel estimado de resistencia a un determinado ARV, excepto las MRD-IP-otras (**Tabla 2.1**). Así, un virus será susceptible a un IP concreto cuando el *score* total de las MDR-IP que presenta sea de 0 a 9. La resistencia será potencialmente baja cuando el *score* total sea de 10 a 14, baja con *score* total de 15 a 29, intermedia con 30 a 59 y resistencia muy alta con *scores* ≥ 60 (<https://hivdb.stanford.edu/page/release-notes/>).

Para identificar los cambios de aa en Gag relacionados con resistencia a IP, se revisaron un total de 14 publicaciones relevantes: Bennett *et al.*, 2009; Larrouy *et al.*, 2010; Ghosn *et al.*, 2011; Larrouy *et al.*, 2011; McKinnon *et al.*, 2011; Fun *et al.*, 2012; Li *et al.*, 2013; Liegeois *et al.*, 2013; Flynn *et al.*, 2015; Giendhari *et al.*, 2015; Sutherland *et al.*, 2015b; Sutherland *et al.*, 2015a; Giendhari *et al.*, 2016; Rhee *et al.*, 2016.

Una vez identificadas las MRD-IP en Gag y PR, se evaluó la frecuencia de cada una de ellas en cada variante del VIH-1, estableciendo un código de colores para indicar el porcentaje de secuencias de cada variante que presentaba de manera natural la MRD-IP específica: rojo ($\geq 90\%$), naranja (51-89%), azul (10-50%), verde ($<10\%$) y blanco (sin mutación).

Tabla 2.1. Grado de resistencia o *score* previsto por Stanford para cada MRD en la PR Primarias (A), Secundarias (B) y Combinaciones (C)

(A) Primarias							
MRD-IP	IP y <i>scores</i>						
	IDV/r	SQV/r	LPV/r	FPV/r	ATV/r	TPV/r	DRV/r
V32 I	I (15)	I (0)	I (15)	I (30)	I (15)	I (5)	I (15)
M46 I/L	I (10)	I (10)	I (10)	I (10)	I (10)	I (5)	I (0)
	L (10)	L (10)	L (10)	L (10)	L (10)	L (10)	L (0)
I47 V/A	V (15)	V (0)	V (15)	V (35)	V (0)	V (30)	V (10)
	A (15)	A (0)	A (60)	A (60)	A (0)	A (30)	A (10)
G48 V/T	V (10)	V (60)	V (10)	V (0)	V (30)	V (0)	V (0)
	T (10)	T (60)	T (10)	T (0)	T (10)	T (0)	T (0)
I50 L/V	L (-5)	L (-5)	L (-10)	L (-5)	L (60)	L (-5)	L (-10)
	V (0)	V (15)	V (30)	V (60)	V (0)	V (-5)	V (20)
I54 A/L/M/S/T/V	A (15)	A (15)	A (15)	A (10)	A (15)	A (20)	A (0)
	L (10)	L (15)	L (20)	L (60)	L (15)	L (-10)	L (20)
	M (15)	M (15)	M (20)	M (60)	M (15)	M (20)	M (20)
	S (15)	S (15)	S (15)	S (10)	S (15)	S (20)	S (0)
	T (15)	T (15)	T (15)	T (10)	T (15)	T (20)	T (0)
	V (15)	V (15)	V (15)	V (10)	V (15)	V (20)	V (0)
L76 V	V (30)	V (0)	V (30)	V (60)	V (0)	V (-5)	V (20)
V82 A/C/F/L/M/T/S	A (30)	A (15)	A (30)	A (15)	A (15)	A (0)	A (0)
	C (15)	C (15)	C (15)	C (15)	C (15)	C (10)	C (0)
	F (30)	F (10)	F (30)	F (30)	F (15)	F (0)	F (15)
	L (10)	L (10)	L (10)	L (15)	L (10)	L (45)	L (0)
	M (30)	M (15)	M (25)	M (15)	M (10)	M (10)	M (0)
	T (30)	T (15)	T (30)	T (15)	T (30)	T (45)	T (0)
	S (30)	S (15)	S (30)	S (15)	S (30)	S (30)	S (0)
I84 A/C/V	A (60)	A (60)	A (60)	A (60)	A (60)	A (60)	A (30)
	C (60)	C (60)	C (30)	C (60)	C (60)	C (30)	C (15)
	V (60)	V (60)	V (30)	V (60)	V (60)	V (30)	V (15)
N88 S/T/G	S (15)	S (15)	S (0)	S (-10)	S (60)	S (0)	S (-5)
	T (0)	T (0)	T (0)	T (0)	T (15)	T (0)	T (0)
	G (0)	G (0)	G (0)	G (0)	G (15)	G (0)	G (0)
L90 M	M(30)	M(45)	M(15)	M(20)	M(25)	M(0)	M(0)

IP, inhibidores de la proteasa. Los IP están colocados por orden respecto a la fecha de autorización de comercialización (**Anexo II**): IDV: indinavir; SQV: saquinavir; LPV: lopinavir; FPV: fosamprenavir; ATV: atazanavir; TPV: tipranavir; DRV: darunavir. (/r): IP potenciados con ritonavir. LPV es el único IP potenciado con ritonavir comercializado actualmente. DRV y ATV se encuentran coformulados con el potenciador cobicistat (/c) (<http://www.ema.europa.eu/ema/>).

(B) Secundarias							
MRD-IP	IP y scores						
	IDV/r	SQV/r	LPV/r	FPV/r	ATV/r	TPV/r	DRV/r
L10 F	F (10)	F (0)	F (5)	F (15)	F (0)	F (0)	F (5)
K20 T	T (5)	T (5)	T (0)	T (5)	T (5)	T (0)	T (0)
L23 I	I (0)	I (0)	I (0)	I (0)	I (0)	I (0)	I (0)
L24 F/I/M	F (5)	F (5)	F (5)	F (5)	F (5)	F (0)	F (0)
	I (15)	I (10)	I (10)	I (10)	I (10)	I (-5)	I (0)
	M (5)	M (5)	M (5)	M (5)	M (5)	M (0)	M (0)
L33 F	F (5)	F (5)	F (5)	F (10)	F (5)	F (10)	F (5)
K43 T	T (0)	T (0)	T (0)	T (0)	T (0)	T (10)	T (0)
M46 V	V (10)	V (5)	V (5)	V (10)	V (10)	V (5)	V (0)
F53 L	L (0)	L (15)	L (0)	L (0)	L (10)	L (0)	L (0)
Q58 E	E (0)	E (0)	E (0)	E (0)	E (0)	E (15)	E (0)
G73 A/C/D/S/T/V	A (15)	A (15)	A (5)	A (10)	A (10)	A (0)	A (0)
	C (15)	C (15)	C (5)	C (10)	C (10)	C (0)	C (0)
	D (5)	D (10)	D (5)	D (5)	D (5)	D (0)	D (0)
	S (15)	S (15)	S (5)	S (10)	S (10)	S (0)	S (0)
	T (15)	T (15)	T (5)	T (10)	T (10)	T (0)	T (0)
	V (5)	V (10)	V (5)	V (5)	V (5)	V (0)	V (0)
T74 P	P (10)	P (10)	P (5)	P (10)	P (10)	P (25)	P (5)
N83 D	D (10)	D (10)	D (0)	D (0)	D (10)	D (25)	D (0)
N88 D	D (0)	D (10)	D (0)	D (0)	D (10)	D (0)	D (0)
L89 V	V (5)	V (0)	V (0)	V (10)	V (0)	V (0)	V (5)

(C) Combinaciones							
MRD-IP	IP y scores						
	IDV/r	SQV/r	LPV/r	FPV/r	ATV/r	TPV/r	DRV/r
V11 I/L + V32 I	5	0	5	5	0	0	5
V11 I/L + I54 L/M	5	0	5	5	0	0	5
V32 I + I47 A/V	5	0	5	5	0	0	5
V32 I + I54 L/M	5	0	5	5	0	0	5
V32 I + L76 V	5	0	5	5	0	0	5
V32 I + I84 V	5	0	5	5	0	0	5
V32 I + L89 V	5	0	5	5	0	0	5
M46 I/L + I84 V + L90 M	5	5	5	5	5	5	0
M46 I/L/V + L76 V	10	0	10	10	0	0	0
M46 I/L/V + V82 A/C/F/L/M/S/T	10	10	10	10	10	0	0
M46 I/L/V + L90 M	10	5	0	10	10	0	0
I47 V/A + I54 L/M	5	0	5	5	0	0	5
I47 V/A + I84 V	5	0	5	5	0	0	5
F53 L + L90 M	10	10	0	10	10	0	0
I54 A/L/M/S/T/V + V82 A/C/F/L/M/S/T	10	10	10	10	10	0	0
I54 L/M + L90 M	10	10	5	10	10	0	0
I54 L/M + I84 V	5	0	5	5	0	0	5
I54 L/M + L89 V	5	0	5	5	0	0	5
G73 S/T/C/A/D/V + L90 M	10	10	0	10	10	0	0
V82 A/C/F/L/M/TS + L90 M	10	10	5	10	10	0	0

RESULTADOS

RESIDUOS EN GAG ASOCIADOS A MRD-IP ANALIZADOS

Tras analizar 14 publicaciones relevantes expuestas anteriormente, definimos un total de 55 MRD-IP en Gag (54 cambios de aa y 1 delección de aa), ubicados en 16 residuos de 5 proteínas Gag diferentes (P17, P2, P7, P1 y P6^{gag}). Doce (75%) de los 16 residuos estaban localizados en 4 sitios de procesamiento proteico o CS (CS1 o P17/P24, CS3 o P2/P7, CS4 o P7/P1 y CS5 o P1/P6^{gag}). En más detalle, sólo 8 cambios aparecían en 4 posiciones de Gag fuera de GCS (residuos 62, 76, 79 y 81 de P17). Los 46 cambios restantes más la delección se ubicaban en doce posiciones aminoacídicas de proteínas Gag ubicadas en los 4 CS previamente indicados: aa 128 de P17 en CS1, 373 y 374 de P2 junto con 378 en CS3, 428 y 431 de P7 y 436 y 437 de P1 en CS4 y 449, 451, 452 y 453 de P6 en CS5.

Los residuos y cambios específicos en Gag asociados a MRD-IP, su implicación en unión a fármacos, susceptibilidad o fracaso a IP, los previamente descritos como polimorfismos naturales y fuentes de referencia correspondientes se indican en la **Tabla 2.2**.

Tabla 2.2. MRD-IP en Gag considerados y polimorfismos naturales previamente descritos

Proteína Gag (posición)	MRD-IP	Efecto en IP			Polimorfismos naturales	Referencias
		Unión a fármacos	Exposición y/o resistencia	Fracaso		
P17 (MA) (1-131)	G62 A/S/E		A/S	A/S	A/E	Li <i>et al.</i> 2013; Sutherland <i>et al.</i> 2015a; Giendhari <i>et al.</i> 2016
	R76 K	K			K	Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Flynn <i>et al.</i> 2015; Rhee <i>et al.</i> 2016
	Y79 F/H		F	F	F/H	Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Sutherland <i>et al.</i> 2015a; Rhee <i>et al.</i> 2016
	T81 A/L	A/L	A		A/L	Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Sutherland <i>et al.</i> 2015a; Rhee <i>et al.</i> 2016
	V128 I/T/A/del/G/S/D		I/T/A/del/G/S		I/A/G/T/S/D	Larrouy <i>et al.</i> 2010; McKinnon <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Flynn <i>et al.</i> 2015
P2 (363-376)	S373 A/T/P/N/G/V/I		P/T/A/Q/N	P/T/A/Q/N	P/T/A/Q/N	Larrouy <i>et al.</i> 2010; Ghosn <i>et al.</i> 2011; McKinnon <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Giendhari <i>et al.</i> 2015
	A374 N/T/S/V/P/H/G		N/T/S/P	N/T/Del	N/T/S/V/P/H/G	Larrouy <i>et al.</i> 2010; Ghosn <i>et al.</i> 2011; McKinnon <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Giendhari <i>et al.</i> 2015; Giendhari <i>et al.</i> 2016
P7 (NC) (377-431)	M378 Q/I/V				Q/I/V	Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Flynn <i>et al.</i> 2015; Giendhari <i>et al.</i> 2015;
	E428 D/K/G	G	D/G	D/K	G	Ghosn <i>et al.</i> 2011; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Fun <i>et al.</i> 2012; Giendhari <i>et al.</i> 2015; Flynn <i>et al.</i> 2015; Giendhari <i>et al.</i> 2016
	A431 V		V	V	V	Larrouy <i>et al.</i> 2010; Larrouy <i>et al.</i> 2011; Ghosn <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Flynn <i>et al.</i> 2015; Sutherland <i>et al.</i> 2015a; Giendhari <i>et al.</i> 2015; Giendhari <i>et al.</i> 2016

Proteína Gag (posición)	MRD-IP	Efecto en IP			Polimorfismos naturales	Referencias
		Unión a fármacos	Exposición y/o resistencia	Fracaso		
P1 (432-447)	K436 E/R		E/R		R	Larrouy <i>et al.</i> 2010; Larrouy <i>et al.</i> 2011; Ghosn <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Flynn <i>et al.</i> 2015
	I437 T/V/L/F		T/V	T/V	V/L/F	Larrouy <i>et al.</i> 2010; Ghosn <i>et al.</i> 2011; Larrouy <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Flynn <i>et al.</i> 2015
P6 (448-499)	L449 F/P/V		F/P/V	F/P/V	P	Larrouy <i>et al.</i> 2010; Ghosn <i>et al.</i> 2011; Larrouy <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Flynn <i>et al.</i> 2015; Giendhari <i>et al.</i> 2015; Sutherland <i>et al.</i> 2015b; Giendhari <i>et al.</i> 2016
	S451 N/G/R/K		N		N/G/R/K	Larrouy <i>et al.</i> 2010; Ghosn <i>et al.</i> 2011; McKinnon <i>et al.</i> 2011; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Flynn <i>et al.</i> 2015; Giendhari <i>et al.</i> 2015; Sutherland <i>et al.</i> 2015b; Giendhari <i>et al.</i> 2016
	R452 S/K		S/K	S/K		Ghosn <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Liegeois <i>et al.</i> 2013; Flynn <i>et al.</i> 2015
	P453 A/L/T/I		A/L/T	A/L/T	A/L/T/I	Larrouy <i>et al.</i> 2010; Ghosn <i>et al.</i> 2011; Fun <i>et al.</i> 2012; Li <i>et al.</i> 2013; Liegeois <i>et al.</i> 2013; Flynn <i>et al.</i> 2015; Sutherland <i>et al.</i> 2015a; Sutherland <i>et al.</i> 2015b; Giendhari <i>et al.</i> 2015; Giendhari <i>et al.</i> 2016

MA, matriz o P17; NC, nucleocápside o P7; aa, aminoácido; IP, inhibidor de proteasa. Influencia del cambio específico asociado a MRD-IP según las referencias indicadas en: unión a fármacos; disminución de la susceptibilidad a IP o resistencia; fracaso virológico asociado. Indicamos cambios de aa previamente identificados en esas publicaciones como polimorfismos naturales en Gag que aparecen de manera natural en virus de pacientes que nunca han recibido TAR.

VARIABILIDAD NATURAL EN POSICIONES DE GAG Y PR ASOCIADAS A MRD-IP

Considerando todos los datos publicados previamente al respecto, estudiamos la variabilidad natural en esas 16 posiciones de Gag y 37 aa de la PR viral donde previamente se identificaron MRD-IP en cada grupo, subtipo y recombinante CRF del VIH-1 (**Figura 2.1**). El objetivo era confirmar los polimorfismos naturales en esos residuos e identificar otros nuevos identificando las variantes específicas en los que cada uno estaba presente en ausencia de IP. Para ello en junio de 2015 se descargaron 2.141 secuencias pertenecientes a 716 pacientes infectados por alguno de los cuatro grupos virales del VIH-1. Entre las 2.141 secuencias, 2.133 eran del grupo M, 4 del grupo P, 2 del grupo O y 2 del grupo N del VIH-1. Las secuencias correspondientes al grupo M incluyeron 1.140 secuencias de seis de los nueve subtipos existentes (A1, B, C, D, F1 y G) y 993 secuencias de 19 CRF con secuencias disponibles en esa región GagPol a fecha de estudio.

La **Figura 2.1** muestra las MRD-IP en Gag (panel A) y en PR (panel B) evaluadas y la variabilidad natural encontrada en las distintas variantes del VIH-1 (grupos, subtipos, CRF) tras analizar los 150 MRD-IP en Gag (54 cambios de aa y 1 delección) y en PR (96 sustituciones) entre las 2.141 secuencias alineadas que se recuperaron de pacientes *naïve* y que estaban depositadas en el GenBank a fecha de estudio. Entre paréntesis se indica el nº de secuencias GagPol descargadas para cada variante VIH-1 (numerador) y nº de pacientes *naïve* a los que pertenecen dichas secuencias (denominador). Con asterisco se indican aquellas posiciones de *gag* implicadas en un CS. Seis posiciones de la PR (10, 20, 46, 74, 82, 88) se encuentran en más de un grupo de MRD (primarias, secundarias o raras).

Cuando no se indica ningún cambio de aa en las celdas coloreadas de la **Figura 2.1**, todos los MRD-IP en dicha posición se encontraron en las secuencias de la variante VIH-1 correspondiente. Sin embargo, las letras dentro de los cuadros coloreados indican los únicos cambios de aa relacionados con MRD-IP que se encontraban en todas las secuencias de la variante VIH-1 indicada. Para algunas variantes se hallaron cambios diferentes que no eran MRD-IP y que, por tanto, no se muestran en la **Figura 2.1**.

La **Figura 2.1.A** especifica las secuencias con cambios en los 55 residuos de Gag asociados MRD-IP estudiados entre las 2.141 secuencias de pacientes *naïve* infectados

por distintas variantes del VIH-1 recuperadas del GenBank. En este estudio se identifica en detalle los grupos, subtipos, y recombinantes CRF específicos que llevan de manera natural dichos cambios asociados a resistencia a IP, alguno de los cuales aparecen en un gran nº de variantes diferentes, como es el caso de las MRD-IP en los residuos 62, 76, 79, 373, 449, y 451.






Llama la atención la mayor presencia de MRD-IP en ciertas posiciones de Gag en los grupos O, P y M y en determinados recombinantes CRF que están ausentes en la mayoría de subtipos puros del VIH-1. Confirmamos que un gran número de cambios en Gag previamente asociados a MRD-IP eran, en realidad, polimorfismos naturales presentes en más del 90% de las secuencias de ciertos grupos, subtipos y recombinantes del VIH-1. También encontramos un elevado número de subtipos y recombinantes CRF del grupo M del VIH-1 con MRD-IP en Gag presentes entre el 51% y 89% de sus secuencias, principalmente en posiciones 62, 76, 79 y 451. Sin embargo, este estudio es meramente descriptivo y no nos permite conocer el efecto de las MRD-IP en *gag* en la susceptibilidad a IP en los pacientes infectados por esas variantes que los incluyen.

La Figura **2.1.B** resume los resultados tras estudiar la variabilidad natural en 37 posiciones de la PR viral previamente asociadas a resistencia a IP de las 2.141 secuencias de pacientes *naïve* descargadas de GenBank, encontrando una gran conservación entre variantes del VIH-1. Así, observamos que sólo 14 de las 36 MRD-IP primarias descritas en 12 posiciones de la PR estaban presentes de manera natural en el VIH-1, aunque en menos del 10% de las secuencias. Encontramos MRD-IP primarias en el aa 46 de la PR en ciertas variantes del grupo M del VIH-1, pero en bajo número de secuencias: M46I en el sub-subtipo F1 (en 33% de sus 3 secuencias), M46I/L en subtipo B (2.6%), CRF01_AE (0.5%) y M46L en subtipo C (en 0,4% secuencias). El subtipo B presentó cambios en un mayor número de posiciones de la PR asociadas a MRD-IP primarias (10) comparado con el subtipo C (5), recombinante CRF01_AE (2) o sub-subtipo F1 (1).

Cuando analizamos la presencia de las 43 sustituciones de aa en 24 posiciones de la PR viral asociadas a resistencia secundaria a IP (MRD-IP secundarias) observamos mayor grado de variabilidad que la observada en las posiciones de MRD-IP primarias, difiriendo entre variantes del VIH-1 (**Figura 2.1.B**). Las mutaciones **M36I/L/V**, **H69K/R** y

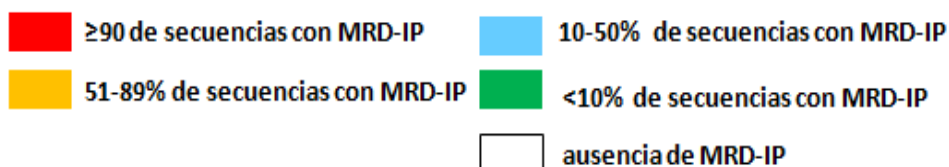
(A)

[illegible]

 ≥90 de secuencias con MRD-IP
  10-50% de secuencias con MRD-IP
 51-89% de secuencias con MRD-IP
  <10% de secuencias con MRD-IP
 ausencia de MRD-IP

(B)

		VIH-1																												
		Grupos no M		Subtipos										Grupo M												CRF				
		N (2/2)	O (2/2)	P (4/2)	A1 (31/18)	B (820/223)	C (281/266)	D (2/2)	F1 (3/3)	G (3/3)	01_AE (920/134)	02_AG (11/4)	07_BC (18/13)	08_BC (2/1)	14_BG (3/3)	15_01B (1/1)	25_cpx (2/2)	26_AU (3/3)	28_BF (3/3)	29_BF (4/4)	33_01B (4/4)	35_AD (9/9)	43_02G (2/2)	47_BF (2/2)	49_cpx (3/3)	52_01B (1/1)	53_01B (1/1)	56_cpx (3/3)	63_02A1 (1/1)	
MRD-IP PRIMARIAS	D30 N																													
	V32 I																													
	M46 I/L																													
	I47 V/A																													
	G48 V/M/ASTQL																													
	I50 L/V																													
	I54 ATS/LM/V																													
	L76 V																													
	V82 A/C/F/L/M/TS																													
	I84 AC/V																													
N88 S/TG																														
L90 M																														
MRD-IP SECUNDARIAS	L10 F																													
	G16 E																													
	K20 T																													
	L23 I																													
	L24 FM/I																													
	L33 F/I/V																													
	E34 Q																													
	M36 I/L/V																													
	K43 T																													
	M46 V																													
	F53 L/Y																													
	Q58 E																													
	D60 E																													
	I62 V																													
	L63 P																													
	I64 L/M/V																													
	H69 K/R																													
	G73 STCADV																													
	T74P																													
	V77 I																													
N83 D																														
N88 D																														
L89 VT/M/I																														
I93 L/M																														
MRD-IP OTRAS	L10 IV																													
	L10 RY																													
	V11 I/ L																													
	K20 I/MV/R																													
	A71 IL/TV																													
	T74 S																													
	V82 I																													
	I85 V																													



También encontramos algunas MRD-IP secundarias con un porcentaje muy variable según la variante VIH-1 de que se trate en varios residuos, como el 16, 33, 58, 60, 62, 63, 64, 77, 89 y 93. Algunas MRD-IP secundarias (L10F/R, L24I, M46V, F53Y/L, G73CSTA, N83D, I85V, y N88D) se encontraron en prevalencias muy bajas (en <1% de las secuencias) en ciertas variantes (**Figura 2.1.B**).

Cuando analizamos otras mutaciones (MRD-IP-otras), denominadas así en el programa HIVdb de Stanford, los cambios **L10R** e **I85V** se encontraron únicamente en un bajo nº de secuencias del subtipo B (0,12 %) y C (0,71%), respectivamente. **K20I** o **K20R** se encontraron en todas las secuencias del grupo N (100%, K20IR), del subtipo G (100%, K20I), en el CRF02_AG (90,9%, K20I), CRF14_BG (100%), CRF25_cpx (100%), CRF26_AU (100%, K20R), CRF35_AD (100%, K20R), CRF43_02G (100%, K20I), CRF52_01B (100%, K20R), CRF56_cpx (100%, K20I) y CRF63_02A1 (100%, K20I) (**Figura 2.1.B**). Algunas de las MRD-IP encontradas no han sido identificadas por Stanford y están descritas por IAS (Wensing *et al.*, 2017): **G16E**, **M36I/L/V**, **D60E**, **I62V**, **L63P**, **I64L/M/V**, **H69K/R**, **V77I**, **N83D**, **I93L/M**. Sin embargo, los cambios encontrados en nuestras secuencias **L10I/V**, **V11I**, **K20R**, **A71V/T**, **V82I**, **T74S** en algunas variantes VIH-1 se señalan como mutaciones polimórficas en Stanford.

El nº de MRD-IP secundarias y raras que aparecían como polimorfismos naturales en $\geq 90\%$ de las secuencias de la variante concreta en ausencia de IP difería entre variantes del VIH-1 (**Tabla 2.3**). Conocer esos datos es importante desde el punto de vista clínico, ya que la acumulación de MRD-IP secundarias se ha asociado a fracaso terapéutico a IP en las variantes con mayor nº (Karmochkine *et al.*, 2000; Luk *et al.*, 2001).

Los grupos O y P del VIH-1 y el recombinante CRF08_BC del grupo M del VIH-1 presentaron el mayor nº de MRD-IP secundarias en la PR y el subtipo B del VIH-1 ninguna. El grupo O mostraba de manera natural los cambios M36I+A71V, previamente asociados a una mayor predisposición a adquirir la MRD-IP primaria D30N (Clemente *et al.*, 2003). Por otro lado, los grupos O y P mostraron el polimorfismo L89I asociado previamente a una mayor resistencia a nelfinavir (NFV) cuando estaba en presencia de L90M respecto cuando a éste último cambio estaba solo (Abecasis *et al.*, 2006). No obstante, NFV ya no se encuentra comercializado en la UE (<http://www.ema.europa.eu/ema/>).

Un hecho importante es que identificamos que 18 (41,9%) de las 43 MRD-IP secundarias y 7 (41,2%) de las 17 MRD-IP “otras” podían considerarse polimorfismos naturales en la PR asociados a la variante viral del VIH-1 concreta, al estar presentes en un 90% o más de las secuencias PR analizadas, de dicha variante en ausencia de tratamiento ARV. Los polimorfismos naturales en posiciones MRD-IP no primarias podrían ser considerados marcadores genéticos de las variantes donde aparecen, pudiéndose encontrar en ese porcentaje en un rango de 1 a 9 según la variante (**Tabla 2.3**).

En más detalle, se podrían considerar como polimorfismos naturales en algunas variantes no-B, al estar presentes en $\geq 90\%$ de sus secuencias, algunas de las MRD-IP secundarias (**G16E, L33I, L33V, M36I, M36L, M36V, Q58E, D60E, I62V, L63P, I64V, L64M, H69K, H69R, V77I, L89I, L89M, L89V e I93L**) y MRD-IP “otras” (**L10I, L10V, V11I, K20I, K20R, A71V y V82I**). Los seis residuos de la PR que presentaban polimorfismos naturales en un mayor número de variantes diferentes al subtipo B del VIH-1 (variantes no-B) en posiciones de MRD-secundarias y raras fueron: 10, 20, 36, 69, 89 y 93. Es de destacar que algunos de los virus de los subtipos G, CRF02_AG, CRF14_BG, CRF25_cpx, CRF43_02G, CRF56_cpx portaban K20I+M36I, quizás debido a que todos esos recombinantes llevaban subtipo G en su PR.

Por el contrario, los cambios K20R+M36I estaban presentes en todos los virus pertenecientes a CRF26_AU, CRF35_AD y CRF52_01B analizados. K20 se mantuvo conservada en todos virus del subtipo D, CRF07_BC y CRF08_BC, CRF15_01B, CRF29_BF, CRF47_BF, CRF49_cpx y CRF53_01B (**Figura 2.1.B**).

Tabla 2.3. Polimorfismos naturales en PR asociados a MRD-IP presentes de manera natural en las distintas variantes del VIH-1

Variante del VIH-1		MRD-IP secundarias (≥90%)	MRD-IP otras (≥90%)	Polimorfismos naturales (≥90%)
Grupo O		7	2	M36I, Q58E, I62V, I64V, H69R, L89I, L10I, A71V
Grupo P		6	3	L33V, M36L, D60E, I62V, H69R, L89I, L10V, V11I, A71V
Grupo N		3	3	L33I, M36IL, H69R, L10V, V11I, K20IR
Grupo M	Sub subtipo A1	3	0	M36I, H69K, L89M
	Subtipo C	2	0	H69K, I93L
	Subtipo G	3	2	M36I, H69K, L89M, K20I, V82I
	Sub subtipo F1	2	0	M36I, L89M
	Subtipo D	1	0	V77I
	Subtipo B	0	0	-
	CRF01_AE	3	0	M36I, H69K, L89M
	CRF02_AG	2	1	M36I, L89M, K20I
	CRF07_BC	2	0	L63P, I93L
	CRF08_BC	6	0	M36V, D60E, L63P, H69K, L89M, I93L
	CRF14_BG	3	2	M36I, H69K, L89M, K20I, V82I
	CRF15_01B	3	0	M36I, H69K, L89M
	CRF25_cpx	4	2	M36I, I64M, H69K, L89M, K20I, V82I
	CRF26_AU	4	1	G16E, M36I, I62V, L89M, K20R
	CRF28_BF	1	0	M36I
	CRF29_BF	2	0	M36I, L89M
	CRF33_01B	3	0	M36I, H69K, L89M
	CRF35_AD	3	1	M36I, H69K, L89M, K20R
	CRF43_02G	3	2	M36I, H69K, L89M, K20I, V82I
	CRF47_BF	2	0	M36I, H69K
	CRF49_cpx	4	0	M36I, H69K, L89M, I93L
	CRF52_01B	3	1	M36I, H69K, L89M, K20R
	CRF53_01B	4	0	G16E, M36I, H69K, L89M
	CRF56_cpx	4	1	M36I, I64V, H69K, L89M, K20I
	CRF63_02A1	5	1	G16E, M36I, I64M, H69K, L89M, K20I

Sólo se indican los que aparecen de manera natural (en ausencia de IP) en $\geq 90\%$ de las secuencias de la variante concreta.

Es importante señalar que ninguna de las 2.141 secuencias GagPol analizadas presentó la MRD-IP primaria L76V ni las MRD-IP secundarias K20T, L23I, E34Q y T74P, pudiendo ser considerados cambios que difícilmente aparecen en pacientes *naïve* y que se seleccionan con el TAR (**Figura 2.1**). Los polimorfismos naturales en residuos asociados a resistencia a ARV en uso terapéutico ubicados en las tres proteínas Pol (PR, RT, IN) y en Gag del VIH-1 asociados a cada grupo, subtipo y CRF del VIH-1 y las MRD que no aparecen en ninguna de nuestras secuencias de pacientes *naïve* analizadas se indican en la **Tabla D.1** y **Tabla D.2** de la Discusión integradora respectivamente. Estos resultados están en la actualidad en fase de preparación para su publicación en una revista internacional de alto impacto.

EFFECTO DE LA EXPOSICIÓN A IP EN LA CONSERVACIÓN DE 11 POSICIONES DE GAG ASOCIADAS A MRD-IP EN MUESTRAS CLÍNICAS

Con el fin de evaluar si la exposición a IP en los pacientes afectaba al grado de conservación de las MRD-IP en *gag*, empleamos un panel de 149 secuencias *gag* recuperadas de virus procedentes de 106 pacientes *naïve* para IP y 43 pacientes con experiencia previa a IP. Los pacientes estaban infectados por distintas variantes del grupo M del VIH-1. Los alineamientos se hicieron con respecto a la secuencia consenso de consensos descargada del GenBank. La región de 515 nt analizada de *gag* incluía 4 CS y 11 residuos de Gag previamente asociados a MRD-IP en 4 proteínas Gag: residuos 373 y 374 en P2, 378, 428 y 431 en P7, 436 y 437 en P1, y 449, 451, 452, y 453 en P6.

DESCRIPCIÓN DE LA POBLACIÓN DE ESTUDIO

La **Tabla 2.4** presenta las características epidemiológicas y clínicas de los 149 pacientes infectados por VIH-1, cuyas secuencias *gag* estaban disponibles en el Laboratorio EpimolVIH para realizar este análisis. La mayor parte (106 pacientes, 71%) de secuencias *gag* procedían de pacientes *naïve* para IP (sin experiencia previa a IP) y las restantes (43 pacientes, 29%) se recuperaron de pacientes con un régimen TAR que incluía IP. Tres cuartas partes de los pacientes (77%) eran de origen extranjero en seguimiento clínico en Madrid, siendo originarios principalmente de África subsahariana,

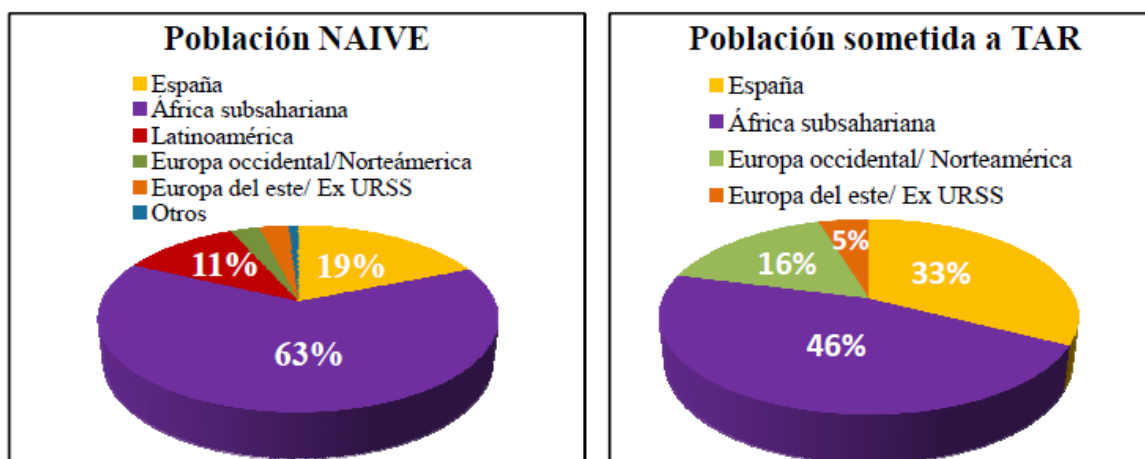
donde las variantes no-B del VIH-1 predominan. El resto (23%) procedían de España (**Figura 2.2**). Cerca de la mitad (48%) eran hombres y la principal vía de transmisión fueron las relaciones homo/bisexuales sin protección (45% de los casos). Cerca de dos tercios (64%) de los 149 pacientes estaban infectados por variantes no-B del VIH-1, previamente caracterizadas por análisis filogenético de las secuencias virales en la región *gag* (Holguín *et al.*, 2005). Encontramos distintos subtipos (18 secuencias pertenecientes al subtipo A, 54 B, 4 C, 2 F, 13 G, 1 H) y recombinantes (45 CRF02_AG, 4 CRF11_cpx, 1 CRF14_BG, 7 URF) del VIH-1.

Tabla 2.4. Descripción de 149 pacientes infectados por VIH-1 con secuencia *gag*

		Total	Naïve	Tratados
		Nº (%)	Nº (%)	Nº (%)
		149	106 (71%)	43 (29%)
Origen	España	34 (23%)	20 (19%)	14 (32.5%)
	Extranjeros	115 (77%)	86 (81%)	29 (67%)
Sexo	Hombre	72 (48%)	47 (44%)	25 (58%)
	Mujer	55 (37%)	40 (38%)	15 (35%)
	Desconocido	22 (15%)	19 (18%)	3 (7%)
Grupo de riesgo	Heterosexual	27 (18%)	20 (19%)	7 (16%)
	Homo/Bisexual	67 (45%)	49 (46%)	18 (42%)
	UDVP	6 (4%)	5 (5%)	1 (2%)
	Vertical	2 (1.3%)	0	2 (5%)
	Transfusión	1 (0.7%)	0	1 (2%)
	Desconocido	46 (31%)	32 (30%)	14 (32,5%)
TAR con IP	Sí	43 (29%)	0	43 (100%)
	No	106 (71%)	106 (100%)	0
Variante del VIH-1	Subtipo B	54 (36%)	33 (31%)	21 (49%)
	Variantes no-B	95 (64%)	73 (69%)	22 (51%)
	Subtipos no-B	38/95 (40%)	27/73 (37%)	11/22 (50%)
	Recombinantes CRF	57/95 (60%)	46/73 (63%)	11/22 (50%)
Periodo de toma de muestra	<2000	18 (12%)	10 (9%)	8 (19%)
	2000-2003	58 (39%)	30 (28%)	28 (65%)
	2004-2006	6 (4%)	6 (6%)	0
	Desconocido	67 (45%)	60 (57%)	7 (16%)

UDVP, usuarios de drogas por vía parenteral; TAR, tratamiento antirretroviral; IP, inhibidores de la proteasa viral; CRF, formas recombinantes circulantes.

Figura 2.2. Origen geográfico de los 106 pacientes *naive* y los 43 tratados con IP



EFFECTO DE LA EXPOSICIÓN A IP EN EL GRADO DE CONSERVACIÓN DE 4 CS

Entre los 4 CS analizados, en los virus recuperados de 149 pacientes con y sin exposición a IP la conservación de aa en los sitios CS4 (P7/P1), CS2 (P24/P2) y CS5 (P1/P6) fueron mayores que la del primer sitio de procesamiento CS3 (P2/P7). Los resultados de conservación observados, considerando sólo las secuencias de los 4 CS de los 106 pacientes *naive*, fueron similares a los obtenidos tras analizar secuencias del GenBank recuperadas de pacientes *naive* representantes de todas las variantes del VIH (**Figura 1.4, capítulo 1 de esta Tesis**). La presencia de IP en esos pacientes parece que no afectó significativamente el grado de conservación de estos sitios, exceptuando en las variantes del subtipo G del grupo M, donde el TAR con IP aumentó de manera significativa la conservación de aa del CS5 (P1/P6gag) del 79% al 97,7% ($p=0,033$) (**Tabla 2.5**). Cuando se consideraron todas las secuencias de los 149 pacientes, la conservación del CS5 tendió a aumentar en virus de pacientes tratados, pero sin llegar a la significación estadística (90,4% vs. 93,5%, $p = 0.054$), lo que sugeriría el interés de ampliar estos estudios a un mayor nº de secuencias en un futuro.

Tabla 2.5. Conservación de 4 CS en virus de distintas variantes en 149 pacientes infectados con y sin experiencia previa a IP

Variante VIH-1	Secuencias		aa conservados/ aa totales		p	aa conservados/ aa totales		p	aa conservados/ aa totales		p	aa conservados/ aa totales		p
	Pac. Naive	Pac Tratados IP	CS2 Naive	CS2 Tratados IP		CS3 Naive	CS3 Tratados IP		CS4 Naive	CS4 Tratados IP		CS5 Naive	CS5 Tratados IP	
A	11	7	110/110 (100%)	68/70 (97,1%)	NS	70/110 (63,6%)	42/70 (60,0%)	NS	105/110 (95,0%)	70/70 (100%)	NS	98/110 (89,1%)	62/70 (88,6%)	NS
B	33	21	323/330 (97,9%)	207/210 (98,6%)	NS	213/330 (64,5%)	137/210 (65,2%)	NS	326/330 (98,8%)	207/210 (98,6%)	NS	317/330 (96,1%)	207/210 (98,6%)	NS
C	3	1	30/30 (100%)	10/10 (100%)	---	19/30 (63,3%)	8/10 (80,0%)	NS	30/30 (100%)	10/10 (100%)	---	27/30 (90,0%)	7/10 (70,0%)	NS
F	2	0	18/20 (90,0%)	—	---	11/20 (55,0%)	—	---	20/20 (100%)	—	---	18/20 (90,0%)	—	---
G	10	3	99/100 (99,0%)	29/30 (96,7%)	NS	55/100 (55,0%)	19/30 (63,3%)	NS	100/100 (100%)	30/30 (100%)	---	79/100 (79,0%)	29/30 (96,7%)	0.024
H	1	0	10/10 (100%)	—	---	7/10 (70,0%)	—	---	10/10 (100%)	—	---	10/10 (100%)	—	---
Total subtipos	60	32	590/600 (98,3%)	314/320 (98,1%)	NS	375/600 (62,5%)	206/320 (64,4%)	NS	591/600 (98,5%)	317/320 (99,1%)	NS	549/600 (91,5%)	305/320 (95,3%)	0.033
CRF02_AG	35	10	349/350 (99,7%)	99/100 (99,0%)	NS	224/350 (64,0%)	66/100 (66,0%)	NS	347/350 (99,1%)	99/100 (99,0%)	NS	310/350 (88,6%)	88/100 (88,0%)	NS
CRF11_cpx	4	0	40/40 (100%)	—	---	27/40 (67,5%)	—	---	39/40 (97,5%)	—	---	37/40 (92,5%)	—	---
CRF14_BG	1	0	10/10 (100%)	—	---	6/10 (60,0%)	—	---	10/10 (100%)	—	---	10/10 (100%)	—	---
URF	6	1	59/60 (98,3%)	10/10 (100%)	NS	45/60 (75,0%)	7/10 (70,0%)	NS	57/60 (95,0%)	10/10 (100%)	NS	52/60 (86,7%)	9/10 (90,0%)	NS
Total recombinantes	46	11	458/460 (99,6%)	109/110 (99,1%)	NS	302/460 (65,7%)	73/110 (66,4%)	NS	453/460 (98,5%)	109/110 (99,1%)	NS	409/460 (88,9%)	97/110 (88,2%)	NS
TOTAL	106	43	1048/1060 (98,9%)	423/430 (98,4%)	NS	677/1060 (63,9%)	279/430 (64,9%)	NS	1044/1060 (98,5%)	426/430 (99,1%)	NS	958/1060 (90,4%)	402/430 (93,5%)	0.054

Leyenda Tabla 2.5. Pac, pacientes con secuencias analizadas; aa, aminoácido. N, naive; IP, tratados con inhibidores de la PR (IP); P24/P2 (CS2); P2/P7 (CS3); P7/P1 (CS4); P1/P6 (CS5). Recomb, recombinantes; p-valores resultantes de comparar las tasas de conservación de aa en las secuencias en los sitios CS analizados tomadas de pacientes *naive* y tratados con IP calculados con STATA; NS, p > 0.05.

DISCUSIÓN

La disponibilidad de nueva información sobre la función de las proteínas Gag, en el ensamblaje, liberación y maduración del VIH-1, ha permitido conocer las vías de la célula huésped que el virus utiliza para promover su replicación y factores de restricción celular que impiden la infección del virus. Sin embargo, el papel de las proteínas Gag apenas ha tenido aún aplicación clínica en el desarrollo de nuevos ARV inhibidores de la maduración (Waheed and Freed, 2012) a pesar de que los IP bloquean el procesamiento de Gag y GagPol por interferir con la actividad enzimática de la PR del VIH-1. Bevirimat es el primer ARV perteneciente a la familia de los inhibidores de la maduración, cuyo mecanismo de acción está relacionado con la función de Gag, aunque actualmente hay más compuestos de esta familia de inhibidores de maduración en desarrollo para su uso clínico (Blair *et al.*, 2009; Titolo 2010; Urano *et al.*, 2015; Spearman 2016). Bevirimat impide la formación de la proteína p24 de la cápsida viral, al inhibir el procesamiento de los precursores virales Gag y GagPol por la PR viral (Li *et al.*, 2006). Sería un ARV dirigido frente al sustrato específico de la PR viral y no frente a la propia enzima viral. El fracaso del tratamiento de los pacientes en un estudio clínico de fase II de bevirimat se atribuyó a polimorfismos naturales en las posiciones de unión a fármacos, apareciendo en patrones subtipo específicos (Adamson *et al.*, 2010; Li *et al.*, 2013). Algunos autores encontraron una asociación entre MRD a bevirimat y MRD a IP, encontrando que algunos polimorfismos frecuentemente observados en los residuos 6-8 del espaciador SP1 conferían diferentes grados de susceptibilidad al bevirimat (Verheyen *et al.*, 2010). Por eso, nuevos compuestos derivados del bevirimat y más potentes están en desarrollo (Urano *et al.*, 2015).

En países con elevados recursos y una experiencia acumulada de 15 años con el TAR, entre el 4,8 y 20% de los nuevos diagnósticos están infectados por variantes del VIH-1 con algún grado de resistencia a los fármacos ARV (Descamps *et al.*, 2010). En pacientes que muestran fracaso a IP se ha observado una acumulación de mutaciones no sólo en sus PR, como era de esperar, sino también en proteínas Gag y en ciertos CS en Gag, sugiriendo un papel en la susceptibilidad a IP (Zhang *et al.*, 1997; de Oliveira *et al.*, 2003; Fun *et al.*, 2012) y en la recuperación de *fitness* reducido previamente por mutaciones de resistencia en la PR (Kaufmann *et al.*, 2001; Nijhuis *et al.*, 2005; Parry *et al.*, 2009). Por tanto, la diversidad genética en Gag puede contribuir a la resistencia a IP, y las

mutaciones en los CS pueden conferir resistencia o susceptibilidad disminuida a IP y emerger antes que las mutaciones en la PR (Lambert-Niclot *et al.* 2012; Fun *et al.* 2012; Giandhari *et al.* 2016). Esto explicaría la ausencia de MRD-IP en la PR en muchos pacientes con TAR incluyendo IP con FV.

Así, las sustituciones observadas en los sitios de procesamiento de *gag* o GCS en los residuos 431, 436 y 437 en el CS4 (NC/p1), 449 en el CS5 (P1/P6), 453 en el CS8 (P6/PR) se han asociado con una disminución en la susceptibilidad a IP y FV durante el tratamiento con IP en ausencia de sustituciones en la PR (Giandhari *et al.* 2016). También se relacionó el polimorfismo natural P453L en *gag* con MRD a IP vía I84V vs. V82I (Bally *et al.*, 2000). Otros estudios han propuesto distintas asociaciones entre ciertos cambios en *gag* con MRD-IP: A431V (en NC) asociado a MRD a IP I54V y V82A, I437V (en P1) con I54V-V82F/T/S y L449F/R452S/P453L (en P6) con D30N-I84V (Verheyen *et al.* 2006). Se ha descrito que virus con las mutaciones L76V en la PR o I437T/V en *gag* podrían inducir resistencia a darunavir (DRV) sin necesidad de existir ninguna otra mutación (Lambert-Niclot *et al.*, 2008). Incluso existen estudios que demuestran que ciertas mutaciones en *gag* aceleran la aparición de MR a ciertos IP, pero retrasan la adquisición de resistencias para otros IP (Aoki *et al.*, 2009). Nuestro análisis realizado en la primera parte de este capítulo ha identificado cambios de aa en todos estos residuos en un gran número de variantes, con una mayor prevalencia en variantes recombinantes CRF y en grupos no-M del VIH-1. No hemos encontrado la asociación MRD-IP en Gag A431V (en NC) asociado a MRD-IP I54V y V82A en PR, ni tampoco I437V (en P1) con I54V-V82F/T/S (en PR), ni L449F/R452S/P453L (en P6) con D30N-I84V (en PR).

También se ha observado que cambios en *gag* cooperan con mutaciones en PR originando resistencia a IP. Así, en un análisis longitudinal de pacientes en FV de primera línea con IP incluyendo lopinavir (LPV), la mutación L76V en la PR, previamente asociada a fracaso con LPV, se asociaba con la mutación M46I en la PR y con el cambio A431V en *gag* (Knops *et al.*, 2010) localizada en el sitio de procesamiento CS4 (P7/P1). Este cambio en *gag* ya había sido descrito previa y exclusivamente en el 23% de pacientes con FV a IP (Bally *et al.*, 2000). Nuestro estudio ha identificado la presencia natural del cambio A431V solamente en un mínimo nº de cepas de subtipo B (0,98%) y subtipo C (0,4%) de pacientes *naive*, pero en ninguna del resto de variantes del VIH-1 analizadas.

Actualmente tan solo un número limitado de estudios evalúa la variación en *gag* y en CS entre variantes (de Oliveira *et al.*, 2003; Fun *et al.*, 2012) y la mayoría de los datos están centrados en el subtipo B o en pocas variantes no-B, siendo también limitado el número de secuencias incluidas de cada uno. En pocos estudios se ve la influencia del TAR (Bally *et al.*, 2000). Previamente se había descrito que el TAR no afectaba de manera significativa ni al grado de inserciones ni de deleciones en algunas proteínas Gag como P6 (Holguín *et al.*, 2005), pero no se había estudiado cómo el TAR con IP puede afectar al grado de conservación de los CS en las distintas variantes del VIH-1. Por ello, en la segunda parte de este capítulo, se evaluó la presencia de cambios en 4 GCS previamente asociadas a mutaciones de resistencia a IP y también cómo el TAR con IP afecta al grado de conservación en los CS en distintas variantes del grupo M del VIH-1. Nuestros resultados han identificado una variante (subtipo G) donde el TAR con IP aumentó significativamente el grado de conservación de uno de los CS (CS5, P1/P6^{gag}) tras comparar secuencias del CS5 de virus de subtipo G de pacientes *naïve* y tratados con IP. No sabemos si una mayor conservación de ese sitio podría tener alterada la capacidad proteolítica del corte por la PR en las variantes del subtipo G, o si los polimorfismos naturales presentes en la PR de subtipo G podrían verse compensados por esa mayor conservación del CS6. Tampoco sabemos si la mayor variabilidad (menor conservación de aa) presente en algunos CS como el CS3 (P2/P7) se pueda ver compensada por mutaciones en la PR específicas de variante. Se desconoce si aumentando el tamaño muestral de esas variantes analizadas, algunas de las diferencias de conservación en los GCS de virus entre pacientes *naïve* y tratados con IP, se pudiera hacer significativa.

La primera limitación del estudio expuesto en este capítulo es que no incluye secuencias de todas las variantes del VIH-1 y que el nº de secuencias por variante no es homogéneo entre los grupos estudiados, tanto en secuencias del GenBank como en las secuencias de los 149 pacientes infectados. En el primer caso se debe a que descargamos las secuencias de GagPol que tenían los genes codificantes de ambas proteínas del mismo paciente en una misma secuencia, con el fin de analizar posible coevolución de MRD Gag y Pol. A pesar de ello, es el trabajo que mayor nº de CRF diferentes incluye. En el segundo caso, a pesar del bajo nº de variantes diferentes, el % de variantes no-B entre los 149 pacientes con VIH-1 de estudio fue seis veces superior a la tasa de variantes no-B en España (64% vs. 10%), al haber priorizado aquellos pacientes que procedían de zonas

endémicas para infecciones por cepas no-B, como es África subsahariana. Por ello, a pesar de que el número de secuencias *gag* asociado a cada variante es pequeño, para ciertos variantes no habría sido posible encontrar el mismo número de todas ellas en pacientes infectados en España (Yebra et al., 2012).

La segunda limitación es que no hemos podido evaluar la influencia de las MRD-IP en Gag en la susceptibilidad en IP, ya que el algoritmo de Stanford sólo permite incorporar secuencias Pol. Ello impide evaluar el impacto de los polimorfismos Gag en la resistencia a IP en la práctica clínica de pacientes en TAR con IP.

Una tercera limitación es que el estudio es descriptivo y no se incluyen estudios funcionales. Estos nos habrían dado más información acerca del efecto de la mayor o menor conservación en cada CS y variante del VIH-1 en la susceptibilidad a fármacos y en la maduración del virus y de la función específica en el ciclo viral de los residuos conservados en las 149 variantes analizadas en posiciones MRD-IP en *gag* y PR. También se requerirían estudios funcionales que expliquen el efecto de una mayor conservación en la eficacia de corte de la PR viral y si la PR de cada variante coevoluciona con Gag para garantizar una eficacia óptima de corte en los sitios más variables. La complejidad elevada, coste de estos ensayos y la ausencia de tiempo real para llevarlos a cabo justifican que no se hayan incluido.

En este capítulo hemos visto que las distintas variantes del VIH-1 presentan distinto grado de conservación en motivos que deberían de estar conservados para garantizar un buen procesamiento y maduración del VIH-1. Se desconoce la repercusión biológica de esas diferencias en la conservación de los aa observadas en los CS y el efecto que tendría en el procesamiento y en el ciclo viral en los distintos grupos, subtipos y recombinantes del VIH-1 (Lee *et al.*, 2012). Por otro lado, puesto que la región *gag* generalmente no se secuencía cuando se producen los fracasos terapéuticos durante el seguimiento clínico de los pacientes con VIH en TAR con IP y, que Stanford no tiene disponible el algoritmo que prediga resistencia a IP en base a MRD-IP en *gag*, resulta difícil identificar las MRD-IP en Gag y entender bien el alcance en la respuesta terapéutica de su asociación con las MRD-IP durante el seguimiento clínico de los pacientes infectados.

CAPÍTULO 3

Variabilidad natural en posiciones de la RT asociadas a resistencia a inhibidores de la transcriptasa inversa en las distintas variantes del VIH-1 en pacientes *naïve*

“Uno de los indicios más fuertes de que hay vida inteligente en el espacio extraterrestre es que no se ha hecho intento alguno de establecer contacto con nosotros”

Normandy Alden

JUSTIFICACIÓN Y OBJETIVOS

La transcriptasa inversa (TI) o retrotranscriptasa (RT) es una proteína viral heterodímera formada por las proteínas p66 y p51. Se encarga de la transcripción inversa del ARN genómico viral y formación del ADN complementario de doble cadena que se integrará en el genoma de la célula infectada (**Paso 3 Figura 1.3 de Introducción**). Los fármacos inhibidores de la transcriptasa inversa (TI), análogos de nucleósido y nucleótido (ITIAN) junto con los no análogos de nucleósido (ITINAN) integran una de las principales estrategias del TAR combinado. La combinación de dos fármacos ITIAN y un ITINAN constituye una de las pautas de TAR combinado y con determinados fármacos puede emplearse como TAR de primera línea.

Los fármacos inhibidores de la TI fueron los que primero se desarrollaron para el tratamiento del VIH-1, muchas veces administrados en dosis y regímenes subóptimos de monoterapia y biterapia. Ello ha contribuido, junto con la baja barrera genética de muchos de ellos, a un frecuente desarrollo de resistencias. Los virus de pacientes con un prolongado uso de ITIAN a menudo acumulan mutaciones que confieren resistencia específica a los diferentes fármacos usados. Es frecuente que el tratamiento con uno o dos ITIAN puede desembocar en la selección de virus con mutaciones de multirresistencia a casi todos los ITIAN disponibles (Jiamsakul *et al.*, 2014).

Se conoce que las mutaciones de resistencia a fármacos en el gen *pol* del VIH-1 representan uno de los factores críticos asociados con el fracaso de la terapia ARV en pacientes infectados por este virus (Misbah *et al.* 2016). En el caso concreto de las mutaciones de resistencia a los ITINAN cabe destacar su elevada frecuencia y el aumento de su prevalencia en los últimos años en algunas zonas geográficas, reflejando los patrones locales del uso de ARV (Rhee *et al.*, 2015).

Como ya hemos ido comentando a lo largo de esta Tesis, durante la evolución viral se fijan cambios diferenciales en las distintas variantes del VIH (polimorfismos naturales) a lo largo de su genoma que se podrían considerar como marcadores genéticos de los diferentes tipos, grupos, subtipos y recombinantes del VIH-1. Algunos de esos

polimorfismos se localizan en posiciones de proteínas donde se han localizado cambios asociados a resistencias a ARV y fracaso terapéutico. Por ello, la disponibilidad de secuencias genéticas en GenBank en el gen de la RT permite estudiar la variabilidad natural en esas posiciones asociadas a resistencia a ARV de la familia de los ITIAN e ITINAN. Sin embargo, la mayor parte de los estudios sobre mutaciones genéticas en las proteínas del VIH-1 seleccionadas como dianas farmacológicas para el TAR se limitan a pacientes con VIH-1 subtipo B, y no se han realizado esfuerzos suficientes para identificar mutaciones de resistencia a fármacos en una escala más amplia, sobre todo en los nuevos recombinantes que están surgiendo en la pandemia. Por todo ello nos propusimos en este trabajo identificar y describir los polimorfismos naturales a nivel de aa de la RT asociados a resistencia a los inhibidores de la TI de la familia de los ITIAN e ITINAN en las distintas variantes del VIH-1 recuperadas de pacientes *naïve* depositadas en el GenBank a fecha de estudio.

Como es conocido, la eficacia del TAR puede verse reducida en pacientes que han seleccionado un virus resistente a ARV tras el TAR (resistencia adquirida). Sin embargo, también pacientes *naïve* a TAR pueden responder peor al tratamiento si han adquirido la infección con una cepa resistente (resistencia primaria), hecho que está aumentando en África, EE.UU., Europa y difiriendo según el origen y el grupo de riesgo de los pacientes (Yebra *et al.*, 2010; de Mulder *et al.*, 2012; Frenz *et al.*, 2014; Yebra *et al.*, 2014; Pham *et al.*, 2014, Rojas *et al.*, 2014, Rhee *et al.*, 2015; Ssemwanga *et al.*, 2015; Margot *et al.*, 2017). A medida que el número de individuos en TAR con mutaciones de resistencia adquirida (MRA) aumenta, también lo hace la proporción de individuos recién infectados con mutaciones de resistencia transmitida (MRT). La instauración de regímenes que incluyen ARV con alta barrera genética a desarrollar resistencia, combinado con una mejora de adherencia al TAR, podrían mitigar el aumento de resistencias en los pacientes infectados. Gracias a ello, en países de Europa Occidental, la emergencia de resistencias a ARV ha disminuido mucho, y la transmisión de cepas resistentes se ha estabilizado en los últimos años. Sin embargo, en muchos países de recursos limitados, aún no se ha regulado las tasas de resistencias transmitidas ni adquiridas (Jiamsakul *et al.*, 2014; Hofstra *et al.*, 2017).

Las tasas de MRT son mayores para ITIAN en Europa (Hofstra *et al.*, 2016), y en muchas partes del mundo llegando incluso a valores del 45% en ciertas regiones (Hofstra *et al.*, 2017). No obstante, el impacto en resistencia es mayor para ITINAN que para ITIAN. Así, un estudio que incluía 50.870 pacientes *naive* de 111 países observó que sólo un pequeño nº de MRD a ITINAN (4 MRD: K101E, K103N, Y181C, y G190A) causaban más del 80% de infecciones en pacientes *naive* en todas las regiones y subtipos (Rhee *et al.*, 2015).

En concreto en España la Red de Investigación en SIDA-CoRIS, en el periodo 2007-2012 estimó una prevalencia global de resistencias primarias en RT y PR del 7,6% (GESIDA 2017; Álvarez M *et al.*, 2015). Durante 1993-2010, la prevalencia global de MRT en la Cohorte de niños y adolescentes infectados por VIH de Madrid, integrado en la Cohorte Nacional de pacientes pediátricos con infección VIH (CoRISPe), fue del 13,6%, mayor que en la Cohorte de adultos de la Red de SIDA (CoRIS) durante 2007-2010 (8,6%) o 2011-2013 (6,9%) o 2007-2012 (7,6%). En la cohorte pediátrica de Madrid las infecciones por virus con triple resistencia a IP, ITIAN e ITINAN fueron frecuentes entre 2000-2011.

Ya que identificar qué MRD-ITIAN y MRD-ITINAN están presentes de manera natural en las distintas variantes del VIH-1 podría ayudar a entender mejor las resistencias a inhibidores de la RT, nos propusimos:

- Identificar mutaciones de resistencia (MRD) a los fármacos inhibidores de la RT del VIH-1 en las distintas variantes del virus con secuencias depositadas en GenBank procedentes de pacientes *naive*.
- Analizar los polimorfismos naturales a nivel de aa de la RT asociados a resistencia a ITIAN e ITINAN, en distintas variantes del VIH-1 recuperadas de pacientes *naive* y depositadas en GenBank.

MATERIAL Y MÉTODOS

SECUENCIAS RT DE PACIENTES NAIVE DESCARGADAS DE GENBANK E IDENTIFICACIÓN DE AA RELACIONADOS CON RESISTENCIA

Para llevar a cabo los objetivos del presente capítulo, en primer lugar en agosto de 2016 recuperamos del GenBank todas las secuencias de la región genómica p51 (RT)

pertenecientes a los grupos O, N, y P, así como de los 9 subtipos y todos los CRF del grupo M del VIH-1 disponibles recuperados de pacientes que no habían recibido previamente TAR o *naive*. La selección de secuencias del VIH-1 pertenecientes a pacientes *naive* de la base de datos del GenBank, se realizó utilizando el filtro “*only drug naive sequences*” y para seleccionar únicamente una secuencia por paciente incluimos el filtro “*one sequence/patient*”. Asimismo, la herramienta “*compose sequence labels*” nos permitió disponer de los datos identificativos de cada individuo con secuencia descargada: subtipo, nº de acceso del GenBank, código de paciente, año de muestreo, país, entre otros.

En la búsqueda de las secuencias p51 de la RT pertenecientes a pacientes *naive* entre las distintas variantes, observamos que algunas variantes del VIH-1 no disponían de secuencias “*drug naive*”, estando representadas estas variantes en GenBank únicamente con secuencias pertenecientes a pacientes en TAR. Para cerciorarnos que no hubiera secuencias de estas variantes pertenecientes a pacientes *naive* en otras regiones genómicas distintas a p51 y que incluyeran la RT, ampliamos nuestra búsqueda en Pol CDS (incluye PR, RT e IN) y el genoma completo. Finalmente contamos para poder llevar a cabo nuestros objetivos con un total de 3.592 secuencias descargadas del GenBank pertenecientes a 3.592 pacientes *naive* a TAR. Con ellas pudimos evaluar la variabilidad asociada a cada variante viral en posiciones de la proteína RT asociadas a susceptibilidad disminuida o resistencia a fármacos ITIAN e ITINAN.

Una vez descargadas y alineadas todas las secuencias, procedimos a estudiar la presencia de los cambios a nivel de aa de la proteína RT asociados a resistencia (MRD-RT) descritos en diversas fuentes de información.

En concreto, se tomaron como referencia las mutaciones descritas por el grupo IAS-USA en 2017 (Wensing *et al.*, 2017), por la base de datos de resistencias a fármacos VIH de la Universidad de Stanford (HIV Drug Resistance Database, <https://hivdb.stanford.edu/dr-summary/resistance-notes/NRTI/>, <https://hivdb.stanford.edu/dr-summary/comments/NRTI/>, <https://hivdb.stanford.edu/dr-summary/resistance-notes/NNRTI/>, <https://hivdb.stanford.edu/dr-summary/comments/NNRTI/>) y también las descritas por Bennett *et al.*, 2009 que recogen las MRT en pacientes *naive* a TAR.

En base a toda la información encontrada, las MRD-ITIAN se clasificaron según muestra la **Tabla 3.1** en 3 grupos:

- A. **Las mutaciones *major* o principales** incluyeron las mutaciones de resistencia múltiple o multiresistencia a ITIAN, representadas por las TAM (mutaciones no polimórficas seleccionadas por los análogos de timidina, AZT y d4T), por el complejo de inserción 69 y por el complejo 151. También incluyeron las MRD discriminatorias, que inducen resistencia mediante un mecanismo discriminatorio o por disminución de unión al sustrato.
- B. **Las mutaciones accesorias** incluían las TAMs accesorias, mutaciones No-TAM, y Otras mutaciones a las que Stanford no ha asignado un *score* de resistencia hasta el momento. En estas Otras mutaciones a menudo son mutaciones que están asociadas con resistencia a los ARV, pero que son principalmente accesorias, y que son polimórficas (es decir, que con frecuencia se producen incluso en personas no tratadas). Sin embargo a veces se incluyen mutaciones raras no polimórficas seleccionadas por el TAR que no han demostrado contribuir a la resistencia a ITIAN.
- C. **Las mutaciones raras o poco frecuentes.**

Es de destacar que algunas MRD-ITIAN se encontraban en dos grupos diferentes según la clasificación de Stanford. Este es el caso de K70/Q/T/N/S, L74V/I e Y115F incluidas en las MRD *major* discriminatorias y MRD accesorias No-TAMs adicionales.

En total, pudimos identificar un total de 69 MRD, 4 delecciones y una inserción ubicadas en 25 residuos aminoacídicos de la RT relacionadas con resistencia a ITIAN y 67 MRD-ITINAN en 25 aa diferentes de la RT (**Tabla 3.1**). Por tanto, las 136 MRD a inhibidores de la RT estaban ubicadas en 50 residuos de la RT diferentes. Observamos varias posiciones de la RT cuyos cambios según su naturaleza, se incluían en distintos grupos. Para ITIAN, eran 11 posiciones de la RT (41, 65, 67, 69, 70, 74, 75, 115, 151, 215, 219) y para ITINAN eran 12 (100, 101, 103, 106, 138, 179, 181, 188, 190, 227, 230, 238).

Figura 3.1. Clasificación de las MRD en la RT a ITIAN e ITINAN

MRD-ITIAN PRINCIPALES	Multi-Resistencia ITIAN	TAMs	M41 L
			D67 N
			K70 R
			L210 W
			T215 Y/F
			K219 Q/E
		C 69	T69 Inserción
		C 151	A62 V
			V75 I
			F77 L
			F116 Y
			Q151 M
		Discriminatorias	K65 R
			K70 E
			L74 V/I
			Y115 F
			M184 V/I
MRD-ITIAN ACCESORIAS		TAMs Accesorias	E40 F
			E44 A/D
			D67 G/E/S/T/H
			V118 I
			T215 S/C/D/E/I/V/A/L/N
			K219 N/R
		No-TAMs Adicionales	K70 E/G/Q/T/N/S
			L74 V/I
			Y115 F
		Miscelanea	T69 D/N/G
			V75 T/M/A/S
		Otras	M41 I
			T69 S/A/I/E
			K219W
MRD-ITIAN RARAS		K65 E	
		K65 N	
		D67 Delection	
		S68Delection	
		T69 Delection	
		K70 Delection	
		K70 Q/T/N/S	
		Q151 L	
MRD-ITINAN PRINCIPALES		L100 I	
		K101 E/P	
		K103 N/S	
		V106 A/M	
		E138 A/G/K/Q	
		Y181 C/I/V	
		Y188 L/C/H	
		G190 A/S/E	
		M230 L	
MRD-ITINAN ACCESORIAS		V90 I	
		A98 G	
		K101 N/A/T	
		K101 Q	
		K103 R	
		V106 I	
		V108 I	
		I132 M/L	
		E138 A	
		E138 Q/G	
		V179 D/E	
		V179 F	
		V179 I	
		H221 Y	
		P225 H	
MRD-ITINAN RARAS		F227 L	
		K238 T/N	
		K238 R	
		N348 I	
		L100 V	
		K103 E/Q	
		K103 H	
		K103 T	
		E138 R	
		V179 L	
		V179 T	
		Y181 F/S/G	
		Y188 F	
		G190 E/Q	
		G190 C/T/V	
		F227 C	
		M230 I	
		P236 L	
		Y318 F	

MRD-ITIAN, mutaciones de resistencia a inhibidores de la transcriptasa inversa análogos de nucleósido o nucleótido; TAMs, mutaciones asociadas a los análogos de la timidina; C 69, complejo de mutaciones de la inserción 69; C 151, complejo de mutaciones 151. Las mutaciones sombreadas en gris se encuentran en más de un grupo dentro de la clasificación. Las mutaciones M41L, A62V, K70R, L210W, T215YF, K219QE se incluyen también en el complejo de mutaciones de la inserción 69. MRD-ITINAN, mutaciones de resistencia a inhibidores de la transcriptasa inversa no análogos de nucleósido.

INTERPRETACIÓN DE LA RESISTENCIA A NIVEL DE GENOTIPO EMPLEANDO EL ALGORITMO HIVdb DE STANFORD

Las secuencias se introdujeron en la aplicación de Stanford (<http://sierra2.stanford.edu/sierra/servlet/JSierra?action=sequenceInput>), para confirmar las MRD-ITIAN e MRD-ITINAN encontradas e interpretar el nivel de resistencia estimado para cada ARV según su puntuación o “*scores*” de mutación: susceptible (0-9), nivel de resistencia potencialmente bajo (10-14), bajo (15-29), intermedio (30-59) o alto nivel de resistencia (≥ 60) (**Tabla 3.2**).

Algunos cambios analizados no se incluyeron en la Tabla porque no tenían *score* asignado por Stanford, como V118 I, V75T/M/A/S, M41I e T69 S/A/I/E (MRD-ITIAN) o V179T y V179I (MRD-ITINAN).

Tabla 3.1. Grado de resistencia o score para ITIAN (A) o ITINAN (B) predicho por Stanford para cada MRD en la RT

MRD-ITIAN (A)	TIPO	ITIAN y score						
		AZT	D4T	3TC	ABC	DDI	TDF	FTC
M41 L	Principales	L (15)	L (15)	L (0)	L (5)	L (10)	L (5)	L 0)
A62 V		V(5)	V(5)	V(5)	V(5)	V(5)	V(5)	V(5)
D67 N		N (15)	N (15)	N (0)	N (5)	N (5)	N (5)	N (0)
K65 R		R (-15)	R (60)	R (30)	R (45)	R (60)	R (60)	R (30)
T69 Inserción		60	60	30	60	60	60	30
K70 E		E (-10)	E (15)	E (10)	E (15)	E (15)	E (15)	E (10)
K70 R		R (30)	R (15)	R (0)	R (5)	R (10)	R (5)	R (0)
L74 V/I		V (0)	V (0)	V (0)	V (30)	V (60)	V (0)	V (0)
		I (0)	I (0)	I (0)	I (30)	I (60)	I (5)	I (0)
V75 I		I (5)	I (5)	I (5)	I (5)	I (5)	I (5)	I (5)
F77 L		L (10)	L (10)	L (5)	L (5)	L (10)	L (5)	L (5)
Y115 F		F (0)	F (0)	F (0)	F (60)	F (0)	F (15)	F (0)
F116 Y		Y (10)	Y (10)	Y (5)	Y (10)	Y (10)	Y (5)	Y (5)
Q151 M		M (60)	M (60)	M (15)	M (60)	M (60)	M (15)	M (15)
M184 V/I		V (-10)	V (-10)	V (60)	V (15)	V (10)	V (-10)	V (60)
		I (-10)	I (-10)	I (60)	I (15)	I (10)	I (-10)	I (60)
L210 W		W (15)	W (15)	W (0)	W (5)	W (10)	W (5)	W (0)
T215 Y/F		Y (40)	Y (40)	Y (0)	Y (10)	Y (15)	Y (10)	Y (0)
		F (40)	F (40)	F (0)	F (10)	F (15)	F (10)	F (0)
K219 Q/E		Q (10)	Q (10)	Q (0)	Q (5)	Q (5)	Q (5)	Q (0)
		E (10)	E (10)	E (0)	E (5)	E (5)	E (5)	E (0)

MRD-ITIAN (A)	TIPO	ITIAN y score						
		AZT	D4T	3TC	ABC	DDI	TDF	FTC
D67 G/E/S/T/H	Accesorias	G (10)	G (10)	G (0)	G (5)	G (5)	G (5)	G (0)
		E (10)	E (10)	E (0)	E (5)	E (5)	E (5)	E (0)
		S (10)	S (10)	S (0)	S (5)	S (5)	S (5)	S (0)
		T (10)	T (10)	T (0)	T (5)	T (5)	T (5)	T (0)
		H (10)	H (10)	H (0)	H (5)	H (5)	H (5)	H (0)
T69 D/G		D (0)	D (10)	D (0)	D (0)	D (30)	D (0)	D (0)
		G (5)	G (10)	G (0)	G (10)	G (10)	G (5)	G (0)
L74 V/I		V (0)	V (0)	V (0)	V (30)	V (60)	V (0)	V (0)
		I (0)	I (0)	I (0)	I (30)	I (60)	I (5)	I (0)
V75 T/M/A/S		T (0)	T (60)	T (0)	T (0)	T (30)	T (0)	T (0)
		M (10)	M (30)	M (0)	M (0)	M (15)	M (0)	M (0)
		A (0)	A (30)	A (0)	A (0)	A (15)	A (0)	A (0)
		S (0)	S (30)	S (0)	S (0)	S (15)	S (0)	S (0)
Y115 F		F (0)	F (0)	F (0)	F (60)	F (0)	F (15)	F (0)
T215 S/C/D/E/I/V/A/L/N		S (20)	S (20)	S (0)	S (5)	S (10)	S (5)	S (0)
		C (20)	C (20)	C (0)	C (5)	C (10)	C (5)	C (0)
		D (20)	D (20)	D (0)	D (5)	D (10)	D (5)	D (0)
		E (20)	E (20)	E (0)	E (5)	E (10)	E (5)	E (0)
		I (20)	I (20)	I (0)	I (5)	I (10)	I (5)	I (0)
		V (20)	V (20)	V (0)	V (5)	V (10)	V (5)	V (0)
		A (20)	A (20)	A (0)	A (5)	A (10)	A (5)	A (0)
		L (20)	L (20)	L (0)	L (5)	L (10)	L (5)	L (0)
		N (20)	N (20)	N (0)	N (5)	N (10)	N (5)	N (0)
K219 N/R		N (10)	N (10)	N (0)	N (5)	N (5)	N (5)	N (0)
		R (10)	R (10)	R (0)	R (5)	R (5)	R (5)	R (0)
K219 W		W (10)	W (10)	W (0)	W (0)	W (5)	W (0)	W (0)

MRD-ITIAN (A)	TIPO	ITIAN y score						
		AZT	D4T	3TC	ABC	DDI	TDF	FTC
K65 E	Raras	E (0)	E (10)	E (0)	E (10)	E (10)	E (10)	E (0)
K65 N		N (-10)	N (30)	N (15)	N (30)	N (30)	N (30)	N (15)
D67 Delección		30	30	15	30	30	30	15
S68 Delección		0	30	15	15	30	15	15
T69 Delección		0	30	15	15	30	15	15
K70 E/G/Q/T/N/S		E (-10) G (-10) Q (0) T (0) N (0) S (0)	E (15) G (15) Q (15) T (15) N (15) S (15)	E (10) G (10) Q (10) T (10) N (10) S (10)	E (15) G (15) Q (15) T (15) N (15) S (15)	E (15) G (15) Q (15) T (15) N (15) S (15)	E (15) G (15) Q (15) T (15) N (15) S (15)	E (10) G (10) Q (10) T (10) N (10) S (10)
K70 Delección		0	30	15	15	30	15	15
Q151 L		L (30)	L (30)	L (10)	L (30)	L (30)	L (10)	L (10)

MRD-ITIAN (A)	TIPO	ITIAN y score						
		AZT	D4T	3TC	ABC	DDI	TDF	FTC
Q151 M + M184 V/I	Combinaciones	10	10	0	0	0	10	0
L210 W + T215 Y/F		10	10	5	10	10	10	5
L210 W + T215 S/C/D/E/I/V/A/L/N		5	5	0	5	5	5	0
E40 F + M41 L + L210 W + T215 Y/F		5	5	0	5	5	5	0
M41 L + L210 W		10	10	0	10	10	10	0
M41 L + L210 W + T215 Y/F		0	0	5	5	0	5	5
M41 L + T215 S/C/D/E/I/V/A/L/N		5	5	0	5	5	5	0
M41 L + T215 Y/F		10	10	5	15	10	10	5
M41 L + E44 A/D + L210 W + T215 Y/F		5	5	0	5	5	5	0
M41 L + D67 E/G/N + T215 Y/F		5	5	5	5	5	5	5
K65 R + Q151 M		10	0	0	0	0	10	0
D67 E/G/N + T215 Y/F + K219 Q/E/N/R		5	5	0	5	5	5	0
D67 E/G/N + K70 R + M184 V/I + K219 Q/E/N/R		0	0	0	20	0	0	0
D67 E/G/N + K70 R + K219 Q/E/N/R		10	10	10	10	10	10	10
K70 E/G/Q/T/N/S + M184 V/I		0	10	0	0	0	10	0
K70 R + T215 Y/F		5	5	0	5	5	5	0
L74 V/I + M184 V/I		0	0	0	15	0	0	0
F77 L + F116 Y + Q151 M		10	10	15	10	10	15	15

MRD-ITINAN (B)	TIPO	ITINAN y score			
		NVP	EFV	ETV	RPV
L100 I	Principales	I (60)	I (60)	I (30)	I (60)
K101 E/P		E (30) P (60)	E (15) P (60)	E (15) P (60)	E (45) P (60)
K103 N/S		N (60) S (60)	N (60) S (45)	N (0) S (0)	N (0) S (0)
V106 A/M		A (60) M (60)	A (45) M (60)	A (0) M (0)	A (0) M (0)
E138 A/G/K/Q		A (0) G (10) K (10) Q (10)	A (0) G (10) K (10) Q (10)	A (10) G (10) K (10) Q (10)	A (15) G (15) K (45) Q (15)
Y181 C/I/V		C (60) I (60) V (60)	C (30) I (30) V (30)	C (30) I (60) V (60)	C (45) I (60) V (60)
Y188 L/C/H		L (60) C (60) H (60)	L (60) C (60) H (30)	L (10) C (0) H (0)	L (60) C (0) H (0)
G190 A/S/E		A (60) S (60) E (60)	A (45) S (60) E (60)	A (10) S (10) E (45)	A (15) S (15) E (60)
M230 L		L (60)	L (45)	L (30)	L (60)
A98G	Accesorias	G (30)	G (15)	G (10)	G (15)
V108 I		I (15)	I (10)	I (0)	I (0)
E138A		A (0)	A (0)	A (10)	A (15)
E138Q/G		Q (10) G (10)	Q (10) G (10)	Q (10) G (10)	Q (15) G (15)
V179 D/E		D (10) E (10)	D (10) E (10)	D (10) E (10)	D (10) E (10)
V179F		F (15)	F (10)	F (15)	F (15)
H221 Y		Y (15)	Y (10)	Y (10)	Y (15)
P225 H		H (45)	H (45)	H (0)	H (0)
F227 L		L (30)	L (15)	L (0)	L (0)
K238 T/N		T (30) N (10)	T (30) N (10)	T (0) N (0)	T (0) N (0)
N348 I		I (15)	I (0)	I (0)	I (0)

MRD-ITINAN (B)	TIPO	ITINAN y score			
		NVP	EFV	ETV	RPV
L100 V	Raras	V (30)	V (30)	V (10)	V (15)
K103 H		H (60)	H (60)	H (0)	H (0)
K103 T		T (60)	T (15)	T (0)	T (0)
E138 R		R (10)	R (10)	R (10)	R (15)
V179 L		L (10)	L (10)	L (10)	L (15)
Y181 F/S/G		F (60) S (60) G (60)	F (15) S (15) G (15)	F (15) S (15) G (15)	F (30) S (30) G (30)
Y188 F		F (60)	F (60)	F (0)	F (30)
G190E/Q		E (60) Q (60)	E (60) Q (60)	E (45) Q (45)	E (60) Q (45)
G190 C/T/V		C (60) T (60) V (60)	C (60) T (60) V (60)	C (10) T (10) V (10)	C (10) T (10) V (10)
F227 C		C (45)	C (45)	C (30)	C (45)
M230 I		I (30)	I (15)	I (15)	I (30)
Y318 F		F (30)	F (10)	F (0)	F (0)
K101 E + Y181 C	Combinaciones	5	5	5	0
K101 E + M184 I		0	0	0	15
K101 E + Y188 L		0	0	5	0
K101 E + G190 A		0	0	5	0
K101 E + G190 S		0	0	5	0
K103 R + V179 D		20	20	0	15
V106 A + F227 L		0	15	0	0
E138 K + M184 I		0	0	0	15
V179 F + Y181 C		0	0	15	15
V179 T + Y181 C		0	0	10	10
Y181 C + G190 A/C/S/T/V		0	0	10	10
A98 G + Y181 C		5	5	5	5

ITIAN, Inhibidores de la transcriptasa inversa análogos de nucleósido y análogos de nucleótido. ITINAN, Inhibidores de la transcriptasa inversa no análogos de nucleósido. AZT (zidovudina), D4T (estavudina), 3TC (lamivudina), ABC (abacavir), DDI (didanosina), TDF (tenofovir disoproxil fumarato), FTC (emtricitabina), NVP (nevirapina), EFV (efavirenz), ETV (etravirina), RPV (rilpivirina). Los ITIAN e ITINAN están colocados por orden respecto a la fecha de autorización (**Anexo II**). Las mutaciones sombreadas en gris están presentes en más de un tipo de la clasificación de Stanford.

VARIABILIDAD NATURAL EN POSICIONES DE RT ASOCIADAS A MRD

La frecuencia de cada MRD-ITIAN e MRD-ITINAN presente en cada variante VIH-1 fue evaluada. Se estableció un código de colores que representaba el porcentaje de secuencias que presentaron las MRD-ITIAN o MRD-ITINAN (rojo: $\geq 90\%$, naranja: 51-89%, azul: 10-50%, verde: $<10\%$ y blanco: sin mutación).

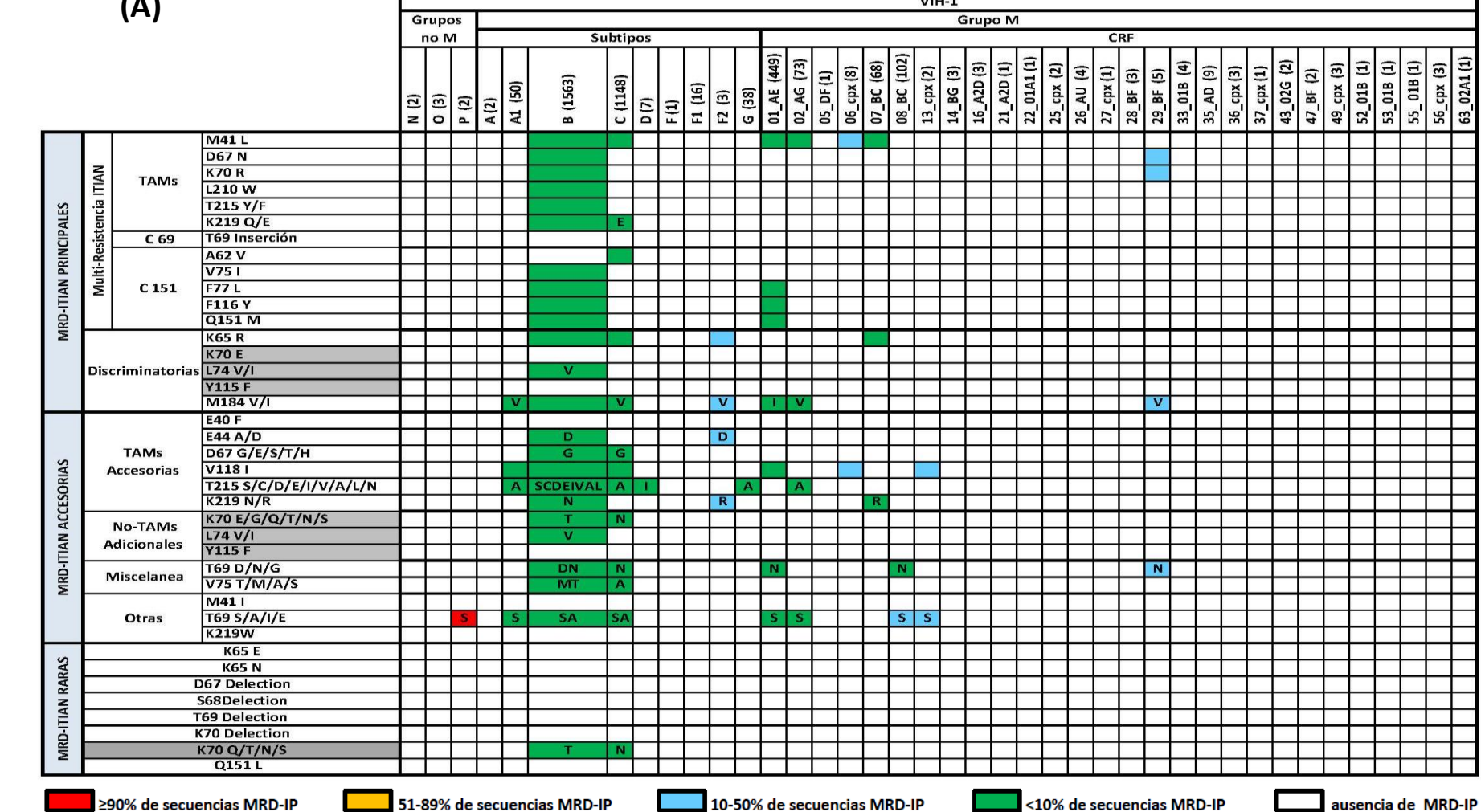
RESULTADOS

VARIABILIDAD NATURAL EN POSICIONES DE LA RT ASOCIADAS A MRD-ITIAN Y MRD-ITINAN EN SECUENCIAS DESCARGADAS DE GENBANK

Para este estudio se descendieron, según se muestra en la **Figura M.1 de Material y Métodos general**, un total de 3.592 secuencias pertenecientes cada una de ellas a un paciente diferente infectado por alguna de las variantes virales del VIH-1. Estas 3.592 secuencias se distribuyen de la siguiente forma: grupo M (3.585), grupo O (3), grupo N (2) y grupo P (2). En concreto las secuencias correspondientes al grupo M incluyeron 2.828 secuencias de los nueve subtipos existentes (2 A, 50 A1, 1.563 B, 1.148 C, 7 D, 1 F, 16 F1, 3 F2 y 38 G) y 757 secuencias de 28 CRF de las 88 CRF descritas en la actualidad.

A partir de las 3.592 secuencias descendidas pudimos estudiar la variabilidad natural en los 50 aa de secuencias de RT donde se han identificado las 136 MRD a ITIAN e ITINAN (**Figura 3.1**). Sólo analizamos la presencia de cambios específicos asociados a MRD mostrados en la **Tabla 3.1** y no de otros aa diferentes, aunque se encontraran. Entre paréntesis en la **Figura 3.1**, indicamos el número de secuencias RT (p51) descendidas para cada variante VIH-1, que coincide con el número de pacientes *naïve* a los que pertenecen dichas secuencias. Es de destacar que el subtipo B presentaba un mayor nº de posiciones de MRD-ITIAN y de MRD-ITINAN afectadas comparado con el resto de variantes, incluyendo otras con un alto nº de secuencias representadas en el estudio, como es el subtipo C y las formas recombinantes CRF01_AE, CRF02_AG, CRF07_BC y CRF08_BC.

(A)	(B)
	XXXX-1



(B)

(B)

		VIH-1																																										
		Grupos no M			Grupo M																																							
		Subtipos										CRF																																
		N (2)	O (3)	P (2)	A (2)	A1 (50)	B (1563)	C (1148)	D (7)	F (1)	F1 (16)	F2 (3)	G (38)	01_AE (449)	02_AG (73)	05_DF (1)	06_cpx (8)	07_BC (68)	08_BC (102)	13_cpx (2)	14_BG (3)	16_A2D (3)	21_A2D (1)	22_01A1 (1)	25_cpx (2)	26_AU (4)	27_cpx (1)	28_BF (3)	29_BF (5)	33_01B (4)	35_AD (9)	36_cpx (3)	37_cpx (1)	43_02G (2)	47_BF (2)	49_cpx (3)	52_01B (1)	53_01B (1)	55_01B (1)	56_cpx (3)	63_02A1 (1)			
MRD-ITINAN PRINCIPALES	L100 I																																											
	K101 E/P						E					E			F																													
	K103 N/S													N	N														N															
	V106 A/M						A	M										M																										
	E138 A/G/K/Q				A		AGK								A	A						A																						
	Y181 C/I/V	C			C	C	C						C		CI			C	C																									
	Y188 L/C/H						LH	L	H						L																													
	G190 A/S/E						AS	A																																				
	M230 L																														A													
MRD-ITINAN ACCESORIAS	V90 I																																											
	A98 G																																											
	K101 N/A/T							N																																				
	K101 Q																																											
	K103 R																																											
	V106 I																																											
	V108 I																																											
	I132 M/L							M																																				
	E138 A																																											
	E138 Q/G							G																																				
	V179 D/E																																											
	V179 F																																											
	V179 I																																											
	H221 Y																																											
	P225 H																																											
	F227 L																																											
	K238 T/N																																											
	K238 R																																											
	N348 I																																											
	MRD-ITINAN RARAS	L100 V																																										
K103 E/Q								Q																																				
K103 H																																												
K103 T																																												
E138 R																																												
V179 L																																												
V179 T																																												
Y181 F/S/G																																												
Y188 F																																												
G190 E/Q																																												
G190 C/T/V																																												
F227 C																																												
M230 I																																												
P236 L																																												
Y318 F																		</																										

MRD-ITIAN

A partir de la **Figura 3.1** podemos interpretar que la presencia de MRD a los ITIAN estudiados fue poco frecuente en aquellos pacientes *naïve* al TAR (**Figura 3.1.A**). Así, las MRD-ITIAN principales o “*major*” aparecían en un bajo porcentaje de secuencias (<10%) en el sub-subtipo A1, en los subtipos B y C, y en los recombinantes CRF01_AE, CRF02_AG y CRF07_BC, pero en mayor frecuencia (10-50%) en el sub-subtipo F2 (2 posiciones) y en los CRF06_cpx (1 posición) y CRF29_BF (3 posiciones) (**Figura 3.1.A**). Entre las **MRD-ITIAN principales a análogos de timidina o TAM, M41L** se encontró en menos del 10% de las secuencias de los subtipos B y C, y de los recombinantes CRF01_AE, CRF02_AG y CRF07_BC, mientras que se localizó con una proporción algo mayor en el CRF06_cpx (entre 10-50% de secuencias). Las TAM **D67N, K70R, L210W y T215Y/F** se localizaron en un bajo porcentaje (<10%) en el subtipo B. Los cambios **D67N** y **K70R** aparecieron en un mayor porcentaje (20%) en el recombinante CRF29_BF. La **mutación en el complejo de inserción 69 K219Q/E** se localizó también en un bajo porcentaje (<10%) en los subtipos B y C. Respecto al **complejo de mutaciones 151**, sólo el subtipo B y recombinante CRF01_AE presentaron 3 de sus mutaciones (**F77L, F116Y y Q151M**) en menos del 10% de sus secuencias y el subtipo C presentó el **A62V** también en baja frecuencia. Respecto a las **MRD-ITIAN principales discriminatorias, K65R** se encontró en menos del 10% de las secuencias de los subtipos B y C y del recombinante CRF07_BC, pero en mayor prevalencia (10-50%) en el sub-subtipo F2. Por su parte, **M184V/I** fue localizada en menos del 10% de secuencias en el sub-subtipo A1, los subtipos B, C y en los recombinantes: CRF01_AE y CRF02_AG en menos del 10%, incrementando su presencia entre 10-50% en el sub-subtipo F2 y CRF29_BF.

La **mutación en el complejo de inserción 59 K219Q/E** se localizó también en un bajo porcentaje (<10%) en los subtipos B y C. Respecto al **complejo de mutaciones 151**, sólo el subtipo B y el recombinante CRF01_AE presentaron 3 de sus mutaciones (**F77L, F116Y y Q151M**) en menos del 10% de sus secuencias y el subtipo C presentó el **A62V** también en baja frecuencia. Respecto a las **MRD-ITIAN principales discriminatorias, K65R** se encontró en menos del 10% de las secuencias de los subtipos B y C y del recombinante CRF07_BC, pero en mayor prevalencia (10-50%) en el sub-subtipo F2. Por su parte, **M184V/I** fue localizada en menos del 10% de secuencias en el sub-subtipo A1, los

subtipos B, C y en los recombinantes: CRF01_AE y CRF02_AG en menos del 10%, incrementando su presencia entre 10-50% en el sub-subtipo F2 y CRF29_BF.

Con respecto a las **MRD-ITIAN accesorias**, **T69S** estuvo presente en las 2 secuencias RT del grupo P analizadas. Encontramos principalmente un bajo grado de variabilidad (MRD-ITIAN accesorias presentes en menos del 10% de secuencias) en el sub-subtipo A1, subtipos B, C, D, G y formas recombinantes CRF01_AE, CRF02_AG, CRF07_BC y CRF08_BC del VIH-1. Algunos otros cambios se encontraron en una proporción mayor (10-50%) de las secuencias del sub-subtipo F2, CRF06_cpx, CRF08_BC, CRF13_cpx y CRF29_BF. En más detalle, el cambio **V118I** fue localizado en pocas secuencias del sub-subtipo A1, subtipos B, C y CRF01_AE (<10% de secuencias), apareciendo con una prevalencia mayor (10-50%) en los recombinantes CRF06_cpx y CRF13_cpx.

La sustitución **K219N** apareció en el subtipo B (<10%), **K219R** en el recombinante CRF07_BC (<10%) y en un mayor porcentaje en el sub-subtipo F2 (10-50%). **E44A/D** en el subtipo B (<10%) y en el sub-subtipo F2 (10-50%). En menos del 10% de secuencias se encontraban las MRD **D67G** en los subtipos B y C, **T69D/N** en subtipo B, **T69N** en subtipo C y recombinantes CRF01_AE y CRF08_BC, **T69 S/A** en subtipos B y C, **T69S** en sub-subtipo A1 y CRF01_AE y CRF02_AG, **K70T** y **K70N** en subtipos B y C, respectivamente, **V75T/M** en subtipo B, **V75A** en subtipo C y **T215S/C/D/E/I/V/A/L/N** en el sub-subtipo A1, subtipos B, C, D, G y CRF02_AG. Una secuencia (20%) del recombinante CRF29_BF presentó la MRD **T69N**. **T69S** apareció en 1 secuencia de los 2 CRF13_cpx (50%) y en 31 de las 102 secuencias CRF08_BC (30%). Respecto a las **MRD-ITIAN raras**, únicamente encontramos la sustitución **K70N** en menos del 10% de las 1.148 secuencias analizadas del subtipo C.

Ninguna de las 3.592 secuencias de la RT del VIH-1 analizadas de pacientes *naive* llevaban ni deleciones en las posiciones 67, 68, 69, 70, ni la inserción en el residuo 69, todas relacionadas con resistencia a ITIAN (ver **Tabla 3.1**). Además ninguna de las secuencias presentó los cambios MRD-ITIAN: **E40F**, **M41I**, **K65E**, **K65N**, **K70E**, **Y115F**, **Q151L** y **K219W**, pudiéndolos considerar como residuos totalmente conservados en nuestro panel de secuencias virales de estudio.

MRD-ITINAN

También la frecuencia natural de las MRD-ITINAN fue poco frecuente en aquellos pacientes *naïve* al TAR (**Figura 3.1.B**). Una excepción es la presencia natural de la MRD-ITINAN principal **Y181C** en los 3 aislados del grupo O y MRD-ITINAN accesorias en los grupos no-M del VIH-1 y CRF35_AD, CRF52_01B, CRF53_01B, CRF55_01B y CRF63_02A1 del grupo M, donde más del 90% de sus secuencias llevaban ciertas MRD-ITINAN principales o accesorias. Algunas MRD-ITINAN se encontraron en una proporción variable (10-50% de sus secuencias) en los grupos N y P, y en el sub-subtipo F2, subtipo D, y formas recombinantes CRF01_AE, CRF06_cpx, CRF08_BC, CRF14_BG, CRF29_BF y CRF33_01B del grupo M del VIH-1. Entre ellos, el CRF29_BF presentó un mayor nº de posiciones MDR-ITINAN afectadas (5) seguida del CRF06_cpx (4), del subtipo D, sub-subtipo F2, CRF14_BG y grupo P (con 3), y CRF08_BC (con 2).

Respecto a las **MRD-ITINAN principales**, **Y181C** estaba asociada al grupo O. Además, 5 variantes del grupo M (subtipo D, sub-subtipo F2, CRF06_cpx, CRF14_BG y CRF29_BF) presentaron entre una y dos MRD-ITINAN principales en un 10-50% de sus secuencias (**Figura 3.1.B**). En otras 7 variantes del grupo M se observó entre 1 y 9 MRD-ITINAN principales pero en un bajo porcentaje de secuencias (<10%), principalmente en aquellas variantes con mayor nº de secuencias analizadas (sub-subtipo A1, subtipos B, C y G y en los recombinantes circulantes: CRF 01_AE, CRF02_AG, CRF07_BC y CRF08_BC del grupo M del VIH-1).

Respecto a las **MRD-ITINAN accesorias**, **A98G** se encontraron en casi la totalidad de los grupos P y O; **V179I** en el grupo N y **V179D/E** en cerca del 100% de los grupos P, O, y en las secuencias de CRF55_01B y CRF63_02A1 del grupo M. Otras MRD-ITINAN accesorias como **V106I**, **V179I** y **K238R** se localizaron en una elevada proporción de secuencias (51-89%) en los sub-subtipos F1, A1, y en el recombinante CRF01_AE, respectivamente.

Varias MRD-ITINAN accesorias (**V90I**, **K101Q/R**, **V106I**, **V108I**, **E138A**, **V179D/E/I**, **H221Y** y **K238T/R**) se encontraron entre un 10-50% de secuencias de 10 variantes del VIH-1 diferentes, entre un rango de 1 y 3 por variante. Sólo 6 variantes del VIH-1 presentaron **MRD-ITINAN raras** y en un rango que oscilaba entre 1 y 7 MRD por variante, presentando el mayor nº el subtipo B del grupo M (**Figura 3.1.B**).

Ninguna de las 3.592 secuencias de la RT del VIH-1 analizadas de pacientes *naive* llevaba los MRD-ITINAN: **K103H/T, E138R, Y181F/S/G, G190C/T/VE/Q, F227C y P236L**.

INTERPRETACIÓN DE LA RESISTENCIA A NIVEL DE GENOTIPO EMPLEANDO EL ALGORITMO HIVdb DE STANFORD

Empleando el programa HIVdb de Stanford con nuestras secuencias analizadas procedentes del GenBank encontramos diferencias en los patrones mutacionales de la RT (**Tabla 3.2**). Sin embargo, al igual que el resultado de nuestro análisis el programa de Stanford tampoco identificó MRD-ITIAN y/o MRD-ITINAN en la RT de nuestras secuencias *pol* VIH-1 para las siguientes variantes VIH-1: Subtipo A, subtipo F, CRF05_DF, CRF16_A2D, CRF21_A2D, CRF22_01A1, CRF25_cpx, CRF26_AU, CRF27_cpx, CRF28_BF, CRF36_cpx, CRF43_02G, CRF47_BF, CRF49_cpx, CRF56_cpx. En CRF13_cpx no se encontraron MRD-ITINAN; mientras que en el grupo N, grupo O, subtipo F1, CRF14_BG, CRF33_01B, CRF35_AD, CRF37_cpx, CRF52_01B, CRF53_01B, CRF55_01B, CRF63_02A1 no se localizaron MRD-ITIAN.

Por otra parte, observamos que para 11 variantes del VIH-1, 9 de ellas variantes recombinantes CRF, la identificación de la variante VIH-1 realizada por Stanford difería de la realizada por GenBank (**Tabla 3.3**).

Tabla 3.2. Diferencias en MRD-ITIAN (A) e MRD-ITINAN (B) empleando el programa HIVdb de Stanford con las secuencias de GenBank

A. MDR-ITIAN

VARIANTE VIH-1	MRD-ITIAN ALGORITMO HIVdb de STANFORD	MRD-ITIAN
Grupo P	-	T69S
Subtipos		
A1	-	M184V, V118I, T215A, T69S
B	-	M41L, D67N, K70R, L210W, T215Y/F, K219Q/E, V75I, F77L, F116Y, Q151M, K65R, L74V, M184V/I, E44D, D67G, V118I, T215SCDEIVAL, K219N, K70T, L74V, T69DN, V75MT, T69SA, K70T
C	-	M41L, K219E, A62V, K65R, M184V, D67G, V118I, T215A, K70N, T69N, V75A, T69SA, K70N
D	T215I	T215I
F2	E44D, K65R, T69Deletion, K219R	K65R, M184V, E44D, K219R
G	-	T215A
CRF		
01_AE	-	M41L, F77L, F116Y, Q151M, M184I, V118I, T69N/S
02_AG	-	M41L, M184V, T215A
06_cpx	M41L, V118I	M41L, V118I
07_BC	-	M41L, K65R, K219R
08_BC	-	T69NS
13_cpx	V118I	V118I, T69S
29_BF	M184V	D67N, K70R, M184V, T69N

B. MDR-ITINAN

VARIANTE VIH-1	MRD-ITINAN ALGORITMO HIVdb de STANFORD	MRD-ITINAN
Grupos		
N	V179I	V106I, V179I
O	A98G, V179E, Y181C	Y181C, A98G, V179E
P	A98G, V179E	A98G, K101Q, K103R, V108I, V179E
Subtipos		
A1	V179I	E138A, Y181C, V106I, E138A, V179I, N348I, V179T
B	-	L100I, K101E, K103NS, V106A, E138AGKQ, Y181C, Y188LH, G190AS, M230L, V90I, A98G, K101Q, K103R, V106I, V108I, E138A, E138QG, V179DEFI, H221Y, P225H, F227L, K238TNR, N348I, L100V, K103EQ, V179LT, Y188F, M230I, Y318F
C	-	K103NS, V106M, E138AGK, Y181C, Y188L, G190A, V90I, A98G, K101N, K101Q, K103R, V108I, I132M, E138AG, V179DEI, K238R, N348I, K103Q, V179T
D	-	Y188H, V179I, H221Y
F1	-	V176I
F2	K101E, Y181L	K101E, Y181C, V108I
G	-	V106I
CRF		
01_AE		K103N, E138A, Y181CI, Y188L, V90I, A98G, K101Q, K103R, V106I, E138A, V179DI, P225H, K238R, V179T, M230I
02_AG		K101E, K103N, E138A, V90I, I132M, E138A, V179EI, P236L
06_cpx	H221Y	Y181C, V179EI, H221Y
07_BC	-	V106M, Y181C, K101Q, V179DEI
08_BC	-	E138A, V90I, V106I, E138A, V179DEI, K238R, N348I, V179T
14_BG	V179E	E138A, V179E
29_BF	G190A	K103N, G190A, K101Q, V108I, K238T
33_01B	-	K238R
35_AD	V179I	V179I
37_cpx	Y318L	-
52_01B	K238R	K238R
53_01B	K238R	K238R
55_01B	V179E	V179E
63_02A1	V179E	V179E, K238R

Tabla 3.3. Diferencias en la identificación de la variante VIH-1 de GenBank empleando el programa HIVdb de Stanford

VARIANTE VIH GENBANK	VARIANTE VIH STANFORD	Nº SECUENCIAS ANALIZADAS
CRF14_BG	Subtipo G	3
CRF16A2D	Subtipo A2	3
CRF22_01A1; CRF26AU	Subtipo A	1; 4
CRF28BF ; CRF29BF	Subtipo B	3; 5
CRF33_01B	Subtipo B+CRF01_AE	4
CRF52_01B	Subtipo B+CRF01_AE	1
CRF37_cpx	Subtipo A+G	1
CRF53_01B	CRF01_AE	1
CRF63_02A1	Subtipo A+CRF02_AG	1
Subtipo A1	Subtipo A	50
Subtipo F1	Subtipo F	16

DISCUSIÓN

Los cambios a nivel de aa se traducen en líneas generales en un incremento en la variabilidad de las proteínas virales, contribuyendo en el caso que nos ocupa a la susceptibilidad, exposición, resistencia o fracaso terapéutico a los ITIAN e ITINAN, tanto en pacientes *naïve* como en tratados. El VIH-1 acumula rápidamente cambios genéticos o mutaciones, algunos de los cuales hacen que el VIH sea resistente al tratamiento ARV de forma que ha sido capaz de desarrollar resistencia a todos los fármacos disponibles. Los ITINAN tienen una barrera genética baja a la resistencia, por lo que el margen de éxito del TAR en países con recursos limitados (Rhee *et al.*, 2015a). Entre un 10% y un 30% de los individuos que reciben un TAR de primera línea con ITIAN/ITINAN desarrollarán FV en algún momento durante su tratamiento (Barth *et al.*, 2010; McMahon *et al.*, 2013; Aghokeng *et al.*, 2014; Rhee *et al.*, 2015c) y seleccionarán virus resistentes a estas familias de ARV (Gupta *et al.*, 2009; WHO 2013; Stadeli *et al.*, 2013; Aghokeng *et al.*, 2014; Rhee *et al.*, 2015c).

A medida que el nº de individuos con MRA ha aumentado, también lo ha hecho la proporción de individuos recién infectados con MRT (WHO 2013; Frentz *et al.*, 2012; Gupta *et al.*, 2012; Rhee *et al.*, 2015a; Rhee *et al.*, 2015c). Las MRT tienen el potencial de revertir más rápidamente la efectividad de la terapia ARV de primera línea a nivel de población. Las personas con MRT que comienzan la terapia ARV con una barrera genética inferior a la resistencia tienen un mayor riesgo de FV (Kuritzkes *et al.*, 2008; Wittkop *et al.*, 2011; Hamers *et al.*, 2012; Li *et al.*, 2012; Mollan *et al.*, 2012; Chung *et al.*, 2014; Rhee *et al.*, 2015a).

Por todo ello es importante conocer la variabilidad de la RT del VIH-1 porque, como ya hemos expresado en esta Tesis, es la variabilidad de las proteínas virales del VIH-1 que constituyen las principales dianas farmacológicas de los ARV donde reside la capacidad del virus para desarrollar resistencias a los ARV. Para ello no basta con conocer solo aquellas mutaciones que surgen en la RT en respuesta a la selección de fármacos, sino también resulta necesario un mejor entendimiento de por qué otros cambios en la RT no se ven frecuentemente (Dunn *et al.*, 2015) o por qué algunos siempre aparecen en ciertas variantes del VIH-1 (polimorfismos naturales).

En nuestro panel de 3.592 secuencias de la RT del VIH-1 analizadas de pacientes *naïve* y descargadas del GenBank no encontramos ninguno de los cambios K20T, L23I, E34Q, T74P y L76V en la PR (MRD-IP), E40F, M41I, K65E, K65N, K70E, Y115F, Q151L y K219W en la RT (MRD-ITIAN), ni K103H/T, E138R, Y181F/S/G, G190C/T/VE/Q, F227C y P236L en la RT (MRD-ITINAN) ni las G118R, F121Y, P145S, Q146P, Q148EG, V151LA y S230R en la IN (MRD-INI). Tampoco encontramos delecciones ni inserciones en los residuos 67, 68, 69 y 70 de la RT asociados a MRD-ITIAN. La mayoría de las MRD-ITIAN no encontradas han sido descritas como mutaciones raras e incluso extremadamente raras o poco frecuentes (K65N, K65E, Q151L y K219W respectivamente). Sin embargo, M41I es normalmente un artefacto resultante de hipermutación mediada por la proteína APOBEC3G, que puede causar mutaciones adicionales en el genoma del VIH (Dunn *et al.*, 2015) y no por resistencia. E40F es una mutación accesoria no polimórfica seleccionada por AZT y d4T, que normalmente se da en combinación con M41L, L210W y T215Y. M41L sí la encontramos en nuestras secuencias como comentamos en Resultados. K70E/G produce resistencia de bajo nivel a TDF, ABC, DDI y posiblemente a 3TC y FTC e

incrementa la susceptibilidad a AZT. Y115F causa resistencia intermedia a ABC y baja resistencia a TDF. Al igual que sucede con las MRD-ITIAN, casi todas las MRD-ITINAN no localizadas en nuestras secuencias, exceptuando P236L, son mutaciones no polimórficas raras (K103H, E138R, Y181F/S/G, G190C/T/V) extremadamente raras (K103T, F227C) o menos frecuentes como G190Q. K103H y G190Q producen resistencia alta a NVP y EFV. G190Q alta o intermedia resistencia a ETR y RPV. F227C es seleccionada por ETR y RPV y se asocia con alta resistencia a todos los ITINAN. G190E causa alta resistencia a todos los ITINAN. K103T causa resistencia intermedia a NVP y un pequeño efecto en la susceptibilidad a EFV. E138R confiere una reducción en la susceptibilidad a RPV.

Por otra parte, la proporción de individuos *naive*, no expuestos a ARV, con MRT también está aumentando, a pesar de que muchas mutaciones de resistencia a los fármacos ARV disminuyen el *fitness* del VIH y tienden a perderse rápidamente en individuos no expuestos al TAR (Rhee *et al.*, 2015a).

Las mutaciones MRD-ITIAN y MRD-ITINAN encontradas en las secuencias de pacientes *naive* analizadas (**Figura 3.1**) son mutaciones que estando presentes en pacientes *naive* sin exposición previa a TAR confirman la transmisión de un virus resistente según la última lista publicada por la OMS (Bennet *et al.*, 2009). Casi todas las MRT asociadas a resistencia a ITINAN de esta lista han sido localizadas en nuestras secuencias, exceptuando los cambios Y181V, Y188C y G190E. Con una prevalencia <10%: **L100I** en subtipo B, **K101E** en subtipo B, CRF02AG; **K103N/S** en subtipos B y C, **K103N** en CRF01AE, CRF02AG; **V106A** en subtipo B, **V106M** en subtipo C y en CRF07_BC; **V179F** en subtipo B; **Y181C** en sub-subtipo A1, subtipo B y C y en CRF07_BC; **Y188LH** en subtipo B, **Y188L** en subtipo C y CRF01AE, **Y188H** en subtipo D; **G190AS** en subtipo B, **G190A** en subtipo C; **P225H** en subtipo B y CRF 01_AE; **M230L** en subtipo B. Con una prevalencia superior al 10% (10-50%) **K101E** en sub-subtipo F2, **K103N** en CRF29BF, **Y181C** en sub-subtipo F2 y CRF06_cpx; **Y188H** en subtipo D; **G190A** en CRF29_BF; y ≥90% **Y181C** en el grupo O.

En el caso de las MRT asociadas a resistencia a ITIAN del listado de la OMS también las encontramos en nuestras secuencias, exceptuando los cambios D67GE, T69Ins, K70E, L74I, Y115F y K219R: **M41L** con una prevalencia <10% en subtipos B y C, CRF01_AE, CRF02_AG, CRF07_BC, CRF06_cpx (10-50%); **K65R** en subtipo B, C, CRF07_BC,

sub-subtipo F2 (10-50%); **D67N** en subtipo B, CRF29_BF (10-50%); **T69DN** en subtipo B, **T69N** en subtipo C, en CRF01_AE, CRF08_BC y CRF29_BC (10-50%); **K70R** en subtipo B, CRF29_BF (10-50%); **L74V** en subtipo B; **V75MT** en subtipo B; **V75A** en subtipo C; **F77L** en subtipo B, CRF01_AE; **F116Y** en subtipo B, CRF_01AE; **Q151M** en subtipo B, CRF_01AE; **M184V** en sub-subtipo A1, subtipo C, CRF02_AG, sub-subtipo F2 (10-50%), CRF29_BF (10-50%); **M184I** en CRF01_AE; **M184VI** en subtipo B; **T215A** en sub-subtipo A1, subtipo C, G, CRF02_AG; **T215I** en subtipo D; **T215SCDEIVAL** en subtipo B; **K219QE** en subtipo B, **K219E** en subtipo C, **K219N** en subtipo B y **K219R** en CRF07_BC y sub-subtipo F2 (10-50%).

Por último, hemos visto que el uso de algoritmos de interpretación de resistencia genotípica como el de Stanford, al igual que las herramientas automatizadas de subtipaje rápido del VIH-1, presentan limitaciones, especialmente en el caso de las variantes recombinantes CRF y URF. Esto se debe a la variabilidad entre las distintas cepas del virus, por estar basadas en el subtipo B del grupo M del VIH-1, que como ya se ha comentado es la variante prevalente en los países desarrollados (Yebra *et al.*, 2011). Existe una falta de consenso en las mutaciones consideradas como indicadoras de resistencia en RT que podría conducir a la sobreestimación o infraestimación de la resistencia, aunque esto no significa que algoritmos de interpretación como el de Stanford no sean válidos para analizar las formas no-B, al menos en países desarrollados, ya que para la mayoría de variantes se encontró una buena concordancia. Sin embargo, en países de recursos limitados donde son prevalentes las variantes no-B especialmente los CRF y URF, la eficacia de estos algoritmos puede verse muy limitada. El hecho de encontrar estas discrepancias justificaría incluir más muestras de subtipos no B en los estudios y tener en cuenta los estudios de epidemiología molecular para tratar de reducir los potenciales errores.

CAPÍTULO 4

Variabilidad natural en posiciones asociadas a resistencia a inhibidores de la integrasa (INI) en pacientes *naïve* a INI

“El futuro tiene muchos nombres. Para los débiles es lo inalcanzable. Para los temerosos, lo desconocido. Para los valientes es la oportunidad”

Victor Hugo (1802-1885)

JUSTIFICACIÓN Y OBJETIVOS

Cerca del 90% de los 37 millones de infecciones en todo el mundo se deben a subtipos no-B y recombinantes del grupo M del VIH-1. Por ello, es necesario conocer mejor la variabilidad natural asociada a cada grupo, subtipo y recombinante del VIH-1, identificando los polimorfismos naturales específicos de variante a lo largo del genoma. Es de especial interés conocer el aa que aparece de manera natural en los residuos de la integrasa (IN) asociados a resistencia a los fármacos ARV inhibidores de la integrasa (INI) en cada variante del VIH-1 circulante. Estos polimorfismos podrían influir en la barrera genética y/o conducir a la selección de vías específicas de mutaciones de resistencia a INI (MRD-INI), ya que pueden facilitar o retrasar la aparición de variantes VIH-1 totalmente resistentes a INI. Además, como sucede para todos los fármacos ARV, las mutaciones de resistencia en el VIH-1 pueden limitar el uso clínico de los INI e incrementar el riesgo de fracaso del tratamiento antirretroviral (TAR) que incluye estos fármacos.

La disponibilidad de secuencias de la IN del VIH-1 en el GenBank permite la evaluación de la variabilidad natural asociada a cada variante, incluyendo las posiciones de la IN viral donde aparecen las mutaciones de resistencia principales y accesorias a los fármacos INI aprobados: raltegravir (RAL), elvitegravir (EVG) y dolutegravir (DTG). Los INI se unen al centro activo de la IN viral bloqueando la incorporación del ADN viral en el ADN de la célula huésped. Por ello también se les conoce como inhibidores de la transferencia de cadenas de la IN (INSTI, *Integrase Strand Transfer Inhibitor*).

En la actualidad existen pocos estudios que hayan evaluado la variabilidad natural en residuos asociados a resistencia a INI (INI-R) en subtipos no-B y menos aún en las variantes recombinantes circulantes (CRF) del VIH-1. Por ello, el objetivo del presente capítulo fue evaluar la variabilidad asociada en las MRD-INI a RAL, EVG y DTG en las distintas variantes del VIH-1 empleando para ello todas las secuencias de la IN disponibles en el GenBank.

MATERIAL Y MÉTODOS

DESCARGA DE SECUENCIAS DE LA BASE DE DATOS DEL GENBANK

Para realizar este estudio en enero 2015 se descargaron del GenBank todas las secuencias disponibles de la IN (p31) para todas las variantes VIH-1, pertenecientes a pacientes *naïve* al TAR y a pacientes tratados con familias de ARV diferentes a los INI. Por tanto, ninguna de las secuencias empleadas en este estudio procedía de pacientes con experiencia previa a INI.

Para seleccionar las secuencias de pacientes *naïve* a INI, fue necesario establecer dos criterios adicionales de búsqueda en GenBank. El primer criterio de cribado consistió en seleccionar todas las secuencias de la IN anteriores a 2007, por ser el año de autorización del primer INI (RAL) e inicio del primer ensayo clínico con un INI autorizado en Europa y EE.UU. Por tanto, asumimos que esos pacientes eran INI *naïve* para INI, aunque no necesariamente para otras familias de fármacos. El segundo criterio, consistió en seleccionar todas las secuencias IN desde el año 2007 hasta enero 2015, correspondientes a pacientes *naïve* a TAR, es decir, aquellas secuencias pertenecientes a pacientes que no habían sido tratados con ningún ARV durante esos ocho años. Excluimos todas las secuencias IN posteriores al año 2007 sin información relativa al TAR en los pacientes.

Para ejecutar ambos criterios de selección fue preciso emplear varios filtros durante la selección de secuencias de GenBank (<http://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>) (Figura 4.1). Primero seleccionamos la región genómica de interés (p31, integrasa). Después elegimos el año de la muestra: < 2007 para el primer criterio de búsqueda ó >=2007 para el segundo criterio. A continuación seleccionamos la opción de descarga “sólo secuencias *naïve* a fármacos antirretrovirales (ARV)” en el caso de las secuencias descargadas con el filtro de año >=2007. Así se descargaron todas las secuencias de la IN del VIH-1 (grupos, subtipos y recombinantes CRF) disponibles en el GenBank que cumplían los criterios de selección, al pertenecer todas a virus de pacientes *naïve* para INI.

Figura 4.1. Identificación de los criterios de búsqueda de las secuencias IN de estudio

HIV sequence database

DATABASES SEARCH ALIGNMENTS TOOLS PUBLICATIONS GUIDES Search

Sequence Search Interface

Sequence Information

Upload accession file: Seleccionar archivo Ningún archivo seleccionado

Accession number:

Sequence name:

Sequence length:

exact ☒ Sampling year:

Sampling country:

Less sequence information

Sequence id:

Isolate name:

Sample tissue:

☒ Only drug naive sequences

Coreceptor: list field in output
☐ only CCR5
☐ all CCR5
☐ only CXCR4
☐ all CXCR4
☐ only R5X4

Find all sequences for a specific gene or region (HIV-1, SIVcpz and SIVgor)

Genomic region:

Or define start: and end:

☐ Include fragments of minimum length:

Virus:

Subtype:
 CRF14_BG
 CRF15_01B
 CRF16_A2D
 CRF17_BF
 CRF18_cpx

☐ Include recombinants

GB create date(YYYY):

Clone name:

Culture method:

Comment:

Phenotype: list field in output
☐ NSI
☐ SI

Por otro lado, con el fin de aumentar el número de secuencias de la IN del VIH-1 para ciertas variantes poco representadas o ausentes al hacer los dos criterios de búsqueda explicados anteriormente, procedimos a descargar de esas variantes específicas genomas completos de GenBank de pacientes *naive*, incluyéndolas en el análisis al resto de secuencias IN de pacientes *naive* descargadas para llevar a cabo nuestro estudio y este capítulo de la Tesis.

IDENTIFICACIÓN DE MUTACIONES DE RESISTENCIA A INI DESCRITAS POR STANFORD Y POR OTRAS FUENTES

Para identificar las sustituciones MRD-INI asociadas a cada variante del VIH-1 en las secuencias virales de pacientes *naïve* a INI, las secuencias de nt de IN fueron alineadas y traducidas en secuencias de aa con el programa MEGA versión 6.0 (www.megasoftware.net/). Consideramos un total de 35 sustituciones asociadas a MRD-INI. Gran parte de ellas (20 de 35) estaban incluidas en la base de datos de resistencias a fármacos VIH de la Universidad de Stanford (*HIV Drug Resistance Database*, <https://hivdb.stanford.edu/dr-summary/resistance-notes/INSTI/>). A fecha de estudio, Stanford las clasificó en primarias o *Major Primary Mutations* (T66IAK, E92QGV, Y143RCHASKG, S147G, Q148HKR y N155HTS), secundarias o *Major Accessory Mutations* (H51Y, L74M, T97A, E138KA, G140SAC, V151LA, S153YF, G163RK, S230R y R263K) y raras o poco frecuentes o *Rare Primary INSTI Resistance Mutations* (G118R, F121Y, P145S y Q146P), aunque actualmente Stanford las clasifica en principales, accesorias y otras.

En la **Tabla 4.1** se muestran las MRD-INI y los *scores* para cada MRD-INI y fármaco ARV proporcionados por Stanford (<https://hivdb.stanford.edu/dr-summary/mut-scores/INSTI/>). A través de dicho *score* se asigna una puntuación de penalización (*penalty score*) a cada mutación de resistencia en función de su impacto en la resistencia a diferentes fármacos ARV, en base a los datos obtenidos en numerosos estudios publicados. Así una puntuación de 5-29 pronostica un nivel de resistencia bajo, de 30-59 predice resistencia intermedia y por encima de 60 predice un nivel de resistencia alto. Por primera vez en un estudio similar, identificamos y evaluamos la presencia de 15 cambios adicionales (V72I, L74AI, Q95K, L101I, H114Y, S119R, T124A, T125K, A128T, E138T, Q148EG, V151I, E157Q, G193E y I203M) previamente relacionados con MRD-INI y publicados por diferentes fuentes (**Tabla 4.2**).

Tabla 4.1. Mutaciones de resistencia a INI y scores según Stanford

MUTACIONES MDR-INI	Tipo	INI scores		
		Raltegravir	Elvitegravir	Dolutegravir
T66IAK	Primarias o Principales	I (15) A (15) K (60)	I (60) A (60) K (60)	K (10)
E92QGV		Q (30) V (30) G (15)	Q (60) G (60) V (60)	Q (10) E (5) G (5)
Y143RHCASKG		R (60) C (60) H (60) A (60) S (60) K (60) G (60)	R (10) C (10) H (10) A (10) S (10) K (10) G (10)	-
S147G		-	G (60)	-
Q148HRK		H (60) K (60) R (60)	H (60) K (60) R (60)	H (20) K (20) R (20)
N155HTS		H (60) S (30) T (30)	H (60) S (30) T (30)	H (10)
H51Y	Secundarias o Accesorias	Y (15)	Y (15)	Y (10)
L74M		M (10)	M (10)	M (5)
T97A		A (15)	A (10)	A (5)
E138KA		K (15) A (15)	K (15) A (15)	K (10) A (10)
G140SAC		S (30) A (30) C (30)	S (30) A (30) C (30)	S (10) A (10) C (10)
V151LA		L (30) A (15)	L (60) A (30)	L (15)
S153YF		-	Y (15) F (10)	Y (10) F (10)
G163RK		R (15) K (15)	R (15) K (15)	-
S230R		R (15)	R (15)	-
R263K		K (15)	K (30)	K (15)
G118R	Raras	R (30)	R (30)	R (10)
F121Y		Y (30)	Y (30)	Y (5)
P145S		-	S (60)	-
Q146P		-	P (60)	-

INI scores, scores asociados a cada MRD-INI descritos por Stanford (<https://hivdb.stanford.edu/dr-summary/resistance-notes/INSTI/>); (https://hivdb.stanford.edu/pages/POC_INIDRMs_Summary.html). Los scores asociados a cada MRD-INI se muestran como en (<https://hivdb.stanford.edu/dr-summary/mut-scores/INSTI/>).

Tabla 4.2. Mutaciones de resistencia a INI descritas en otras fuentes diferentes de Stanford

Otras MDR-INI	INI afectado	Fuente de información
V72I	RAL	RIS
L74AI	RAL	Myers and Pillay, 2008
Q95K	RAL, EVG	RIS
L101I	DTG	RIS
H114Y	EVG	RIS
S119R	DTG	Carganico <i>et al.</i> , 2014
T124A	DTG	RIS
T125K	EVG	RIS
A128T	EVG	RIS
E138T	DTG	EMA
Q148EG	RAL	ANRS
V151I	RAL, EVG, DTG	RIS, EMA, Myers and Pillay, 2008
E157Q	RAL, EVG, DTG	RIS, ANRS, EMA, Myers and Pillay, 2008; Carganico <i>et al.</i> , 2014
G193E	DTG	RIS, EMA
I203M	RAL	Myers and Pillay, 2008

RAL, raltegravir; EVG, elvitegravir; DTG, dolutegravir. EMA, Agencia Europea de Medicamentos. Los cambios con referencia EMA provienen de la ficha técnica de cada INI, que incluye datos de resistencia proporcionados por los resultados de ensayos clínicos. RIS, Red de Investigación en Sida de España (<http://retic-ris.net/Programa1/WP3.Resistencias.aspx>; 2013); ANRS, Agencia Nacional Francesa para la Investigación del Sida o (<http://www.hivfrenchresistance.org/2014/tab6.html>); EMA, Agencia Europea de Medicamentos (<http://www.ema.europa.eu/ema/>).

Por último, para evaluar la frecuencia de cada cambio de MRD-INI, en cada variante del VIH-1 a partir de sujetos INI-*naive*, fue establecido un código de colores para indicar el porcentaje de secuencias con el cambio MRD-INI correspondiente.

RESULTADOS

En enero de 2015 se descargaron de GenBank, siguiendo los criterios de selección antes mencionados, un total de 6.706 secuencias de la IN VIH-1 de 3.791 pacientes *naive* para INI. Éstas incluían secuencias pertenecientes a los cuatro grupos del VIH-1: M (6.663), O (24), N (15) y P (4). Dentro de las 6.663 secuencias del grupo M del VIH-1 descargadas se incluían 4.599 secuencias de 2.538 sujetos infectados por alguno de los nueve subtipos del grupo M y 2.064 secuencias de 1.216 sujetos INI-*naive* infectados por alguno de los 54 recombinantes CRF diferentes con secuencias disponibles en la IN a fecha de estudio (**Figura 4.2**).

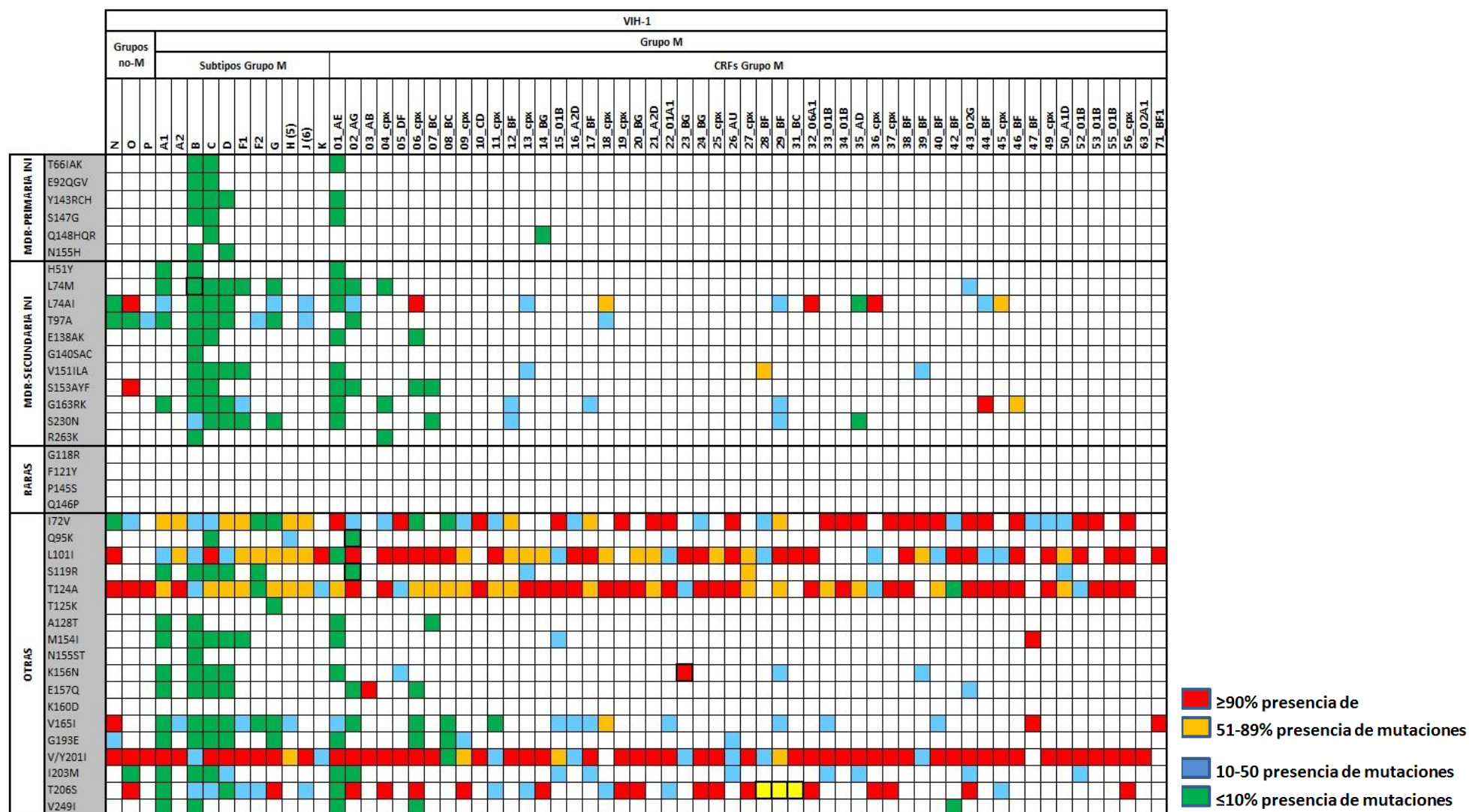
A partir del análisis de la variabilidad en las secuencias descargadas en cada posición asociada a cada una de las 35 MRD-INI descritas en las **Tabla 4.1** y **Tabla 4.2** encontramos que las MRD a INI fueron raras en sujetos INI-*naive* y sólo estuvieron presentes en un bajo porcentaje de secuencias IN de los subtipo B (0,6%), C (0,7%) y D (1,9%) y en los recombinantes CRF01_AE (0,7%) y CRF14_BG (7,1%), presentando de una a cinco mutaciones MRD-INI por variante viral (**Figura 4.2**). Ninguna de las 6.706 secuencias de la IN descargadas de individuos *naive* portaban el cambio MRD-INI principal **Q148EG** o los cambios MRD-INI accesorios **V151LA** y **S230R**. Las cuatro mutaciones MRD-INI raras: **G118R**, **F121Y**, **P145S** y **Q146P** estuvieron también ausentes en todas las variantes estudiadas (**Figura 4.2**), posiblemente se explique por su efecto perjudicial en las propiedades catalíticas de la IN, tal y como se ha publicado para **G118R** y **F121Y** (Munir *et al.*, 2014).

Sin embargo, tres mutaciones MRD-INI secundarias (**L74M**, **T97A** y **G163RK**) aparecieron con una frecuencia variable y en más del 10% de las personas INI-*naive* infectadas por variantes VIH-1 específicas: **L74M** en CRF43_02G (33,3%); **T97A** en grupo P (50%), subtipo J (33,3%), CRF18_cpx (20%) y sub-subtipo F2 (11,5%). **G163RK** presentó alta prevalencia entre algunos recombinantes BF: CRF44_BF (100%), CRF46_BF (66,7%), CRF17_BF (28,6%), sub-subtipo F1 (21,7%), CRF12_BF (16,7%) y CRF29_BF (12,5%). Dada la vital importancia de la respuesta a INI en pacientes infectados con estas variantes virales con MRD-INI, sería fundamental realizar estudios incluyendo un mayor número de muestras para entender mejor su efecto en la susceptibilidad a INI.

Algunas de las 15 mutaciones MRD-INI no incluidas en Stanford pero sí en otras fuentes (**Tabla 4.2**), estuvieron en realidad presentes en $\geq 90\%$ de las secuencias de variantes no-B del grupo M del VIH-1 específicas. En concreto nos referimos a las MRD: **V72I**, **L74AI**, **L101I**, **T124A** y **E157Q**, las cuales podrían ser consideradas como polimorfismos naturales, en un elevado nº de variantes no-B del grupo M (**Figura 4.2**). Por su parte otras MRD altamente presentes fueron: **T124A** que se localizó en todas las secuencias IN de los grupos N y P del VIH-1 y en un 96% de las 24 secuencias del grupo O del VIH-1; **L101I** fue altamente prevalente en el grupo N (93,3%); y **L74I** que apareció en todas las secuencias del grupo O (100%) y en un 6,7% de las secuencias grupo N.

En la **Figura 4.2** la frecuencia de cada cambio de MRD-INI en cada variante del VIH-1 de sujetos INI-*naïve* se indica con un código de colores, representando el porcentaje de secuencias con el cambio MRD-INI correspondiente. En color rojo se destacan las mutaciones encontradas en $\geq 90\%$ de las secuencias IN de la variante del VIH-1 concreta, en amarillo aquellas mutaciones presentes entre 51%-89%, en azul las encontradas entre un 10%-50% y en verde por debajo de un 10%. En blanco se muestran aquellas posiciones donde la MRD-IN no fue encontrada en ninguna de las secuencias de IN analizadas de la variante correspondiente (ausencia de mutación). Las celdas con el borde “en negrita” son cambios en las secuencias IN que no encontramos al analizar las secuencias con el programa HIVdb de Stanford (<https://hivdb.stanford.edu/hivdb/by-mutations/>).

Figura 4.2. Variabilidad natural en cada posición MRD-INI entre las variantes VIH-1



DISCUSIÓN

El punto más significativo a discutir a partir de los resultados obtenidos en este capítulo de Tesis es el hecho de que nuestros datos refuerzan que algunos de los cambios MRD-INI encontrados no incluidos en la base de datos de Stanford, podrían ser polimorfismos asociados a variante más que mutaciones MRD-INI. Sin embargo, consideramos que estos cambios o sustituciones polimórficas deberían ser estudiados con más detalle en un futuro para entender mejor su papel en la evolución de la resistencia a INI y en la vía o mecanismo de resistencia a cada INI en cada variante VIH-1.

Ya se ha publicado que ciertos cambios en posiciones MRD-INI específicas pueden afectar a la actividad nativa, estructura o función de la IN de manera diferente según la variante del VIH-1 (Quashie *et al.*, 2015). Así el impacto de las sustituciones H51Y y E138K en la resistencia y la eficiencia enzimática de la IN in vitro en presencia de la MDR G118R son altamente dependientes del subtipo viral (Quashie *et al.*, 2015). Sin embargo, nuestros datos muestran que esos cambios no son polimorfismos naturales asociadas a variantes concretas. También se ha observado que el coste en eficacia biológica o *fitness* asociado con mutaciones MRD-INI específicas puede ser diferente entre variantes. Por ejemplo, la IN del subtipo C tiene un bajo coste enzimático asociado con la sustitución G118R, comparado con las proteínas IN del subtipo B o del recombinante CRF02_AG (Quashie *et al.*, 2015). La influencia del subtipo del VIH-1 en perfiles mutacionales de la IN han sido recientemente referida en pacientes tratados con RAL (Doyle *et al.*, 2015).

Se sabe que ciertos polimorfismos en la IN presentes de manera natural en determinados subtipos y recombinantes del VIH-1 podrían modificar la barrera genética para la aparición de MRD-INI en esas variantes (Maïga *et al.*, 2009; Turriziani *et al.*, 2012) o inducir vías de resistencia alternativas para MRD-INI (Malet *et al.*, 2011; Quashie *et al.*, 2015), como ya ha sido descrito para algunos inhibidores de la RT ITIAN del VIH-1 (Martínez-Cajas *et al.*, 2009). En cualquier caso, es conveniente resaltar que no todos los polimorfismos naturales en la IN específicos de variante incrementan la resistencia a INI o restauran el coste viral asociado con ciertas mutaciones MRD-INI (Wares *et al.*, 2014).

Nuestro estudio muestra los polimorfismos genéticos naturales presentes en un gran número de variantes del VIH-1, con secuencia disponible en GenBank en sujetos *naive* a los fármacos antirretrovirales inhibidores de la integrasa (INI). Mientras que las mutaciones primarias asociadas a resistencia a los INI (MRD-INI) son poco frecuentes en los individuos INI-*naive* y las MRD-INI raras están ausentes, algunos cambios INI-R secundarios son frecuentes en un número apreciable de variantes de VIH-1. La respuesta a los INI en individuos portadores de estas variantes debe ser mejor estudiada en un futuro para un mejor entendimiento del efecto de estas variantes en la susceptibilidad a los INI.

Nuestros datos también refuerzan que algunos nuevos cambios INI-R encontrados no están incluidos en la base de datos de VIH de Stanford podrían tratarse de polimorfismos asociados a variante más que mutaciones de resistencia a INI.

Los polimorfismos naturales en posiciones asociadas a resistencia a los INI (MRD-INI) que se encuentran en las distintas variantes VIH-1 deberían tenerse en cuenta, ya que puede facilitar o retrasar la aparición de variantes VIH-1 totalmente resistentes a los INI.

Los resultados descritos en este capítulo se recogen en la siguiente publicación:
“Deep analysis of HIV-1 natural variability across HIV-1 variants at residues associated with integrase inhibitor (INI) resistance in INI-naive individuals”. J Antimicrob Chemother.2016; 71(2):362-6. Decil 1. Factor de impacto: 5.4.
<http://www.ncbi.nlm.nih.gov/pubmed/26546669>.

DISCUSIÓN

INTEGRADORA

“Cuando creíamos que teníamos todas las respuestas, de pronto, cambiaron todas las preguntas”

Mario Benedetti (1920-2009)

La variabilidad genética propia de las variantes no-B puede tener implicaciones para prácticamente cada aspecto de la epidemia (**Figura D.1**), por lo que la correcta identificación de las variantes del VIH-1 es de enorme importancia. Los diferentes subtipos del VIH influyen de manera muy amplia y variada en los diferentes y numerosos aspectos de la infección (Geretti 2006; Peeters *et al.*, 2010; Hemelaar 2013). Estos aspectos incluyen el manejo terapéutico (Mackie *et al.*, 2013), manifestaciones clínicas (Lemonovich *et al.*, 2015), progresión más rápida a SIDA, diferente respuesta y/o sensibilidad a determinados fármacos ARV debido a los polimorfismos genéticos (Martínez Cajas *et al.*, 2008), diferente grado de emergencia de mutaciones de resistencia a fármacos (Eshleman *et al.*, 2005; Koning *et al.*, 2013; Menéndez-Arias *et al.*, 2014), y distinto impacto en el sistema nervioso (Rotta *et al.*, 2011; Samikkannu *et al.*, 2014; Samikkannu *et al.*, 2015). Algunos trabajos incluso apuntan a una tasa de transmisión del virus diferente según el subtipo (Kiwana et al., 2009), y ya se ha demostrado el distinto grado de progresión comparando algunas variantes del VIH-1. Así, los pacientes infectados por ciertos subtipos no-B presentan una progresión acelerada de la enfermedad (Vasan *et al.*, 2006; Baeten *et al.*, 2007; Palm *et al.*, 2014; Kouri *et al.*, 2015). Según varios estudios (Pant Pai *et al.*, 2012) los pacientes infectados por el subtipo D presentan, en ausencia de tratamiento, una progresión acelerada a SIDA y un mayor deterioro cognitivo que los infectados por otros subtipos. La recombinación, fenómeno muy frecuente en el VIH como ya se ha comentado, parece ser también un mecanismo clave en su evolución, confiriendo al virus recombinante ventajas evolutivas y adaptativas frente a los virus parentales (Zhang *et al.*, 2010) incluso acelerando la progresión a sida (Palm *et al.*, 2014).

Las técnicas diagnósticas, ya sean directas o indirectas, también pueden verse afectadas por la variabilidad viral (Aghokeng *et al.*, 2009). Los inmunoensayos disponibles actualmente pueden detectar todos los subtipos del grupo M del VIH-1, el grupo O y el VIH-2, aunque en regiones con recursos limitados, donde se concentran la mayor parte de las infecciones y circulan multitud de cepas distintas del virus, las pruebas diagnósticas son menos eficientes (Aghokeng *et al.*, 2009; Plantier *et al.*, 2009).

En cuanto a la detección directa mediante PCR requiere de sondas u oligonucleótidos específicos para amplificar la región diana del virus. Estos, al estar diseñados especialmente para detectar cepas del subtipo B, pueden ser responsables de falsos negativos en el diagnóstico, cuando se trata de detectar subtipos no-B o recombinantes genéticamente diferentes de la cepa B (Díaz Pernas *et al.*, 2006; Bruzzone *et al.*, 2010; Alvarez *et al.*, 2015). La variabilidad genética del VIH también puede ser la responsable de fallos en la cuantificación del VIH, que presentan las técnicas comerciales al determinar la CV en un paciente infectado por variantes no-B (Gottesman *et al.*, 2004; Rouet *et al.*, 2007; Holguín *et al.*, 2008b; Sturdevant *et al.*, 2012; Alvarez *et al.*, 2014, Samikkannu *et al.*, 2015; Tatarelli *et al.*, 2016).

La variabilidad entre las distintas cepas del virus dificulta también el rendimiento de diversas herramientas informáticas como los algoritmos de interpretación de resistencia genotípica y las herramientas automatizadas de subtipaje del virus, durante la monitorización de la infección por VIH, algoritmos en constante evolución (Wagner *et al.*, 2015). Ello se debe a que su diseño está basado principalmente en la detección del subtipo B del VIH-1 (variante mayoritaria en países donde se diseñan los ensayos diagnósticos moleculares), pudiendo provocar errores de interpretación en el diagnóstico de otras variantes, sobre todo recombinantes complejos (Snoeck *et al.*, 2006, Champenois *et al.*, 2008, Poonpiriya *et al.*, 2008; Yebra *et al.*, 2010a; Eberle and Gürtler, 2012).

Figura D.1. Implicaciones diagnósticas y clínicas de las infecciones por variantes no-B del VIH-1



Figura cedida por la Dra. África Holguín de su capítulo: Epidemiología Molecular del VIH. Máster en VIH. Proyecto Esther. 2016 (<https://www.campusesther.org>).

Las pruebas genotípicas para resistencia a ARV emplean técnicas de amplificación por PCR y secuenciación de poblaciones que detectan mutaciones asociadas a resistencia presentes en ≥ 15 -25% de la población viral (Schuurman *et al.*, 1999; Grant *et al.*, 2003; Li *et al.*, 2011). Usando estos ensayos tradicionales, se estima que la prevalencia de mutaciones transmitidas de resistencia a fármacos está entre 8% y 16% entre las personas infectadas por el VIH-1 en América del Norte y Europa (Kim *et al.*, 2010; Vercauteren *et al.*, 2009; Li *et al.*, 2011). Estos ensayos no detectan la presencia de mutaciones de resistencia de baja frecuencia presentes como variantes minoritarias dentro de la población de quasiespecies del VIH-1 en un individuo infectado por lo que las tasas de MRT podrían ser mayores. La presencia de estas variantes minoritarias o de baja frecuencia puede afectar negativamente la respuesta al TAR, pero su importancia clínica continúa siendo objeto de considerable debate e incertidumbre (Li *et al.*, 2011).

La diversidad del VIH es el mayor impedimento para el diseño de una vacuna global y eficaz, sobre todo para su aplicación en zonas donde circulan múltiples cepas del VIH, dada la imposibilidad de desarrollar inmunidad frente a un amplio rango de variantes (Barouch, 2008; Korber y Gnanakaran, 2009, deCamp *et al.*, 2014; Tongo *et al.*, 2014). Las cepas co-circulantes del VIH difieren entre sí un 20% o más en proteínas relativamente conservadas y hasta un 35% en la proteína de la envuelta del virus (Env) (Korber *et al.*, 2001; Gaschen *et al.*, 2002; Fischer *et al.*, 2006). El desarrollo de una vacuna que sea inmunológicamente relevante para múltiples regiones del mundo constituye una prioridad en la investigación (Stephenson and Barouch, 2013; Barouch *et al.*, 2013; Tongo *et al.*, 2014; Excler and Michael, 2016). Para ello, resultaría imprescindible realizar estudios de epidemiología molecular del VIH-1 en zonas concretas para conocer con detalle y rigurosidad la variabilidad geográfica del virus antes del diseño de dichas vacunas (Luo *et al.*, 2013) y nuevos fármacos ARV (Olender *et al.*, 2015). Además, se ha demostrado que la amplitud de las respuestas inmunes celulares específicas de Gag se correlacionan con la replicación del VIH-1 en humanos (Kiepiela *et al.*, 2007; Barouch *et al.*, 2010) por lo que conocer el grado de conservación de estas proteínas entre variantes del VIH resulta esencial.

Respecto a la implicación de la diversidad viral en la eficacia del TAR, la mayoría de los fármacos ARV se dirigen a tres proteínas del VIH-1: PR, RT e IN. Estas proteínas son altamente variables, de modo que muchos aa diferentes pueden estar presentes en la misma posición en los virus de distintos individuos, conformando diferentes variantes genéticas del VIH1. Algunas variaciones o cambios de aa causan resistencia a los fármacos y se producen principalmente en individuos que reciben ARV (Rhee *et al.*, 2016). Como ya hemos comentado a lo largo de esta Tesis, ciertas variantes no-B del VIH-1 poseen de manera natural un alto grado de polimorfismos, presentan sustituciones específicas en posiciones relacionadas con el desarrollo de resistencias a los ARV (Martínez Cajas *et al.*, 2008; Kantor y Katzenstein, 2003) que pueden alterar la aparición de mutaciones de resistencia en esas posiciones o en otras adyacentes, bien acelerando su emergencia (Brenner *et al.*, 2003; Descamps *et al.*, 2006) o bien induciendo nuevas vías alternativas de aparición (Grossman *et al.*, 2004; Gonzalez *et al.*, 2008; Descamps *et al.*, 1998; Brenner *et al.*, 2003). Se han descrito casos en los que ciertas variantes no-B necesitan acumular menos cambios nucleotídicos en las regiones codificantes de PR, RT e IN que el subtipo B para desarrollar una mutación de resistencia, debido a los diferentes tripletes de nt que presentan estas variantes (Martínez-Cajas *et al.*, 2008). Se ha demostrado que la diversidad intersubtipo del VIH-1 afecta al patrón de mutaciones de resistencia a fármacos seleccionadas en el gen *pol* tras el fracaso del tratamiento con los fármacos ARV clásicos dirigidos a la PR y RT (Martínez-Cajas *et al.*, 2009; Rhee *et al.*, 2006; Aulicino *et al.*, 2010; Rossi *et al.*, 2013).

Aunque un nº creciente de estudios observacionales del mecanismo de resistencia a ARV de las variantes no-B ha mostrado que hay diferencias en la evolución de la resistencia entre los subtipos de VIH-1, no se han llevado a cabo esfuerzos para identificar las MRD asociadas a variantes no-B en muchos países donde el VIH-1 constituye una epidemia. En África, sur de Asia incluido India, donde viven la mayoría de personas con VIH, la mayor parte están infectados con el subtipo C (más del 50% de todas las infecciones VIH-1), existe muy poca información sobre la resistencia a ARV en estas personas (Kantor *et al.*, 2005; Misbah *et al.*, 2016). No obstante, el impacto de la variante del VIH-1 (grupo, subtipo, recombinante) en la respuesta clínica al TAR a largo plazo aún es motivo de estudio. Por ello, se necesita continuar el estudio de la eficacia de los ARV

con un mayor número de pacientes infectados por cada variante, comenzando por las más frecuentes. Se debería de hacer más hincapié en las variantes no-B más prevalentes en la pandemia y en los recombinantes cada vez más complejos que están aumentando su prevalencia de manera espectacular en ciertos países donde todas las variantes cocirculan (Ragupathy *et al.*, 2011; Konou *et al.*, 2016; Sharma *et al.*, 2017; Hu *et al.*, 2017; Delatorre *et al.*, 2017). Incluso en España, donde se ha visto un aumento de recombinantes únicos URF entre las nuevas infecciones de niños en los últimos años (Rojas *et al* 2017 *in press*). A pesar de todo esto, la eficacia del tratamiento ARV, produce una buena respuesta virológica e inmunológica en todos los subtipos (Holguín *et al.*, 2006a; Geretti *et al.*, 2009).

El conocimiento de la estructura de las proteínas del VIH-1 y la función de las mutaciones de resistencia a los fármacos de estas proteínas es fundamental para el desarrollo de tratamientos eficaces contra el VIH (Flynn *et al.*, 2015). En los virus resistentes a IP, cambios de aa dentro y fuera de los CS en *gag* han demostrado mejorar su capacidad de replicación y *fitness*, actuando como polimorfismos compensatorios y facilitando el procesamiento de *gag* por una PR mutada (Bally *et al.*, 2000; Verheyen *et al.*, 2006; Rossi *et al.*, 2013). Por ello, un mayor conocimiento de la coevolución de *gag* y *pol* en el VIH-1 (Kozísek *et al*, 2012; Rossi *et al.*, 2013), de la conservación a nivel de aa en los sitios de procesamiento proteico o CS y en la PR (Yebra *et al.*, 2010b) son cruciales para entender mejor el papel biológico de los polimorfismos naturales en los distintos CS y PR presentes en las diferentes variantes del VIH-1 durante el proceso de escisión, aunque se requieran estudios funcionales adicionales para ello. Nuestro estudio descriptivo del Capítulo 1 ha ayudado a avanzar en este conocimiento, ya que ha determinado el grado de conservación de diferentes sitios de procesamiento viral o *cleavage* (CS) en la mayoría de las variantes del VIH-1, a partir de un gran conjunto de datos compuesto por 12.934 secuencias, infiriendo las secuencias consenso a nivel de aa en 11 CS, en todos los subtipos del grupo M y la mayoría de CRF y URF, así como en los grupos O, N y P del VIH-1. Nuestro estudio revela cuáles serían los aa más conservados en los CS (**Figura 1.3**) esenciales para mantener el procesamiento viral por la PR, así como el nivel de tolerancia a un cambio en los aa en cada variante del VIH-1. Los resultados proporcionan nuevos hallazgos que pueden ayudar a una mejor comprensión de la evolución viral y del procesamiento de los precursores Gag y GagPol. Nuestros datos de la

conservación de los CS entre las variantes VIH-1 también podrían resultar útiles para el diseño de fármacos inhibidores de la maduración viral y de vacunas dirigidas a epítomos conservados en los sitios de procesamiento proteolítico que pudieran ser efectivas contra todos los grupos, subtipos y recombinantes de VIH-1. Por último, los resultados obtenidos en este capítulo proporcionan nuevos hallazgos que pueden ayudar a una mejor comprensión de la evolución viral y del procesamiento de los precursores Gag y GagPol. Las observaciones alcanzadas sobre la conservación de los CS entre las variantes VIH-1 también podrían resultar útiles para el diseño de fármacos inhibidores de la maduración viral y de vacunas dirigidas a epítomos conservados en los sitios de procesamiento proteolítico que pudieran ser efectivas contra todos los grupos, subtipos y recombinantes de VIH-1.

Un correcto procesamiento proteico es clave para la conversión de partículas virales no infecciosas inmaduras a viriones infecciosos maduros. La presión selectiva ejercida por los fármacos IP del VIH-1 da lugar a la acumulación de mutaciones de resistencia a estos ARV tanto en la PR como en Gag. Sorprendentemente, un gran nº de estas mutaciones (conocidas como secundarias o compensatorias) se producen fuera del sitio activo de la PR o de los sitios de escisión o CS en Gag, que se han propuesto como cambios esenciales para facilitar el acceso de la PR viral a los sitios de escisión Gag (Deshmukh *et al.*, 2016). Algunos autores observaron un aumento en la selección positiva en los CS de Gag (p7/p24, p24/p2, p2/p7, p7/p1, p1/p6^{gag}) sólo después de la adquisición de mutaciones a IP importantes, lo que sugiere que esos aa bajo presión de selección positiva en esos CS podrían funcionar como mutaciones compensatorias para mutaciones a IP importantes en la región de la PR. Las mutaciones en Gag aisladas no parecían conferir resistencia a IP, pero las mutaciones en los sitios de escisión o procesamiento de Gag podrían sustituir a mutaciones de resistencia a IP secundarias en la región de la PR (Banke *et al.*, 2009).

En la **Tabla D1.2** hemos integrado todos los polimorfismos naturales en Gag, PR, RT e IN asociados a MRD en *gag* y *pol* presentes de manera natural en $\geq 90\%$ de secuencias de las distintas variantes del VIH-1 en ausencia de ARV.

No se encontró ninguna MRD primaria que fuera polimorfismo natural que estuviera presente en 90% o más de las secuencias analizadas de las variantes incluidas, excepto la MRD-ITINAN Y181C en el grupo O y la MRD-ITIAN T69S, ambas en el grupo O. El resto de MRD-RT encontradas como polimorfismos naturales en ciertas variantes del VIH-1 fueron MRD-ITINAN.

Cuando el cambio incluye 2 letras se refiere a que la suma de las secuencias con esas MRD correspondía a $\geq 90\%$ de sus secuencias, poniendo en primer lugar el aa más frecuente. En verde se muestran las variantes del VIH-1 que estaban analizadas en las 4 regiones virales (Gag, PR, RT, IN) pero no en todas. Se indican sombreados en amarillo aquellas MRD encontradas sólo en una variante del VIH-1: K156N en IN de CRF23_BG; M154I en IN de CRF47_BF, E157Q en IN de CRF03_AB y V165I en IN de grupo N. En azul se muestran las MRD consideradas accesorias o secundarias para la familia de ARV correspondiente.

Las variantes que presentaron un mayor nº de MRD (5 o más) fueron los siguientes: en las proteínas Gag (grupos N, O, P y dentro del grupo M los recombinantes CRF15_01B, CRF43_02G, CRF53_01B, CRF63_02A1), en la PR (grupos N, O, P y dentro del grupo M el subtipo G y los recombinantes CRF02_AG, CRF08_BC, CRF14_BG, CRF25_cpx, CRF26_AU, CRF43_02G, CRF56_cpx, CRF63_02A1) y en IN (grupo O y dentro del grupo M los recombinantes CRF32_06A1, CRF43_02G y CRF56_cpx) (**Tabla D.2**).

Las variantes del VIH-1 con un mayor nº de MRD-IP considerando conjuntamente las ubicadas en *gag* y PR, fueron el grupo O y CRF63_02A1 del grupo M (ambas con 13), grupo P y CRF53_01B (ambos con 10), grupo N, CRF15_01B y CRF56_cpx (todos con 9) y CRF08_BC, CRF25_cpx, CRF26_AU y CRF43_02G (todas con 8) (**Tabla D.1**). Se necesitará más investigación de respuesta a IP en pacientes infectados por estas variantes (de baja prevalencia en la pandemia) para entender mejor el efecto de este alto nº de polimorfismos naturales en posiciones asociadas a resistencia a IP.

Tabla D.1. Polimorfismos naturales en Gag, RT, PR e IN asociados a MRD en *gag* y pol presentes de manera natural en $\geq 90\%$ de secuencias de las distintas variantes del VIH-1

Variante del VIH-1	Polimorfismos naturales asociados a MRD (en $\geq 90\%$ de secuencias)				Nº MRD-IP en Gag	Nº MRD-IP en PR		Nº MRD -RT	Nº MRD -INI
	En Gag*	En PR**	En RT***	En IN		secundarias	otras		
GRUPO N	V128I, S373AT, A374T, E428G, L449P, P453T	L33I, M36IL, H69R L10V, V11I, K20IR	V179I	L101I, T124A, V165I, VY201I	6	3	3	1	4
GRUPO O	V128TA, A374T, E428G, L449V, S451RK, P453A	M36I, Q58E, I62V, I64V, H69R, L89I, I93L L10I, A71V	A98G V179E Y181C	L74AI, S153AYF, T124A, VY201I, T206S	6	7	2	3	5
GRUPO P	G62E, E428G, L449V, S451K	L33V, M36L, D60E, I62V, H69R, L89I L10V, V11I, A71V	T69S A98G V179E	T124A, VY201I	4	6	3	3	2
GRUPO M									
Sub subtipo A1	G62E, S373A, L449P	M36I, H69K, L89M	ninguna	VY201I	3	3	0	0	1
Sub subtipo A2	NA	NA	NA	T124A, VY201I	NA	NA	NA	NA	2
Subtipo B	R452K	ninguna	ninguna	ninguna	1	0	0	0	0
Subtipo C	G62ASE	M36ILV, H69K, I93L	ninguna	L101I, VY201I	1	3	0	0	2
Subtipo D	R76K, A374VG	V77I	ninguna	VY201I	2	1	0	0	1
Sub subtipo F1	S373TA	M36I, L89M	ninguna	VY201I	1	2	0	0	1
Sub subtipo F2	NA	NA	ninguna	VY201I	NA	NA	NA	0	1
Subtipo G	R76K	M36I, H69K, L89M K20I, V82I	ninguna	T206S, VY201I	1	3	2	0	2
Subtipo H	NA	NA	NA	ninguna	NA	NA	NA	NA	0
Subtipo J	NA	NA	NA	VY201I	NA	NA	NA	NA	1

Variante del VIH-1	Polimorfismos naturales asociados a MRD (en ≥90% de secuencias)				Nº MRD-IP en Gag	Nº MRD-IP en PR		Nº MRD -RT	Nº MRD -INI
	En Gag*	En PR**	En RT***	En IN		secundarias	otras		
Subtipo K	NA	NA	NA	L101I	NA	NA	NA	NA	1
CRF01_AE	G62EA, S374ATPGVI, L449P	M36ILV, H69K, L89MVI	ninguna	I72V, VY201I	3	3	0	0	2
CRF02_AG	V128del, L449P	M36I, H69RK, L89M L10IV, K20I	ninguna	L101I, T124A, Y201I, T206S	2	3	2	0	4
CRF03_AB	NA	NA	NA	E157Q, VY201I	NA	NA	NA	NA	2
CRF04_cpx	NA	NA	NA	L101I, T124A, VY201I, T206S	NA	NA	NA	NA	4
CRF05_DF	NA	NA	ninguna	I72V, L101I, VY201I	NA	NA	NA	0	3
CRF06_cpx	NA	NA	ninguna	L74AI, L101I, VY201I, T206S	NA	NA	NA	0	4
CRF07_BC	ninguna	L63P, I93L	ninguna	L101I, VY201I	0	2	0	0	2
CRF08_BC	Y79F, I437F	M36V, D60E, L63P, H69K, L89M, I93L	ninguna	L101I	2	6	0	0	1
CRF09_cpx	NA	NA	NA	T206S	NA	NA	NA	NA	1
CRF10_CD	NA	NA	NA	I72V, T124A, VY201I	NA	NA	NA	NA	3
CRF11_cpx	NA	NA	NA	L101I	NA	NA	NA	NA	1
CRF12_BF	NA	NA	NA	VY201I	NA	NA	NA	NA	1
CRF13_cpx	NA	NA	ninguna	T124A, VY201I	NA	NA	NA	0	2
CRF14_BG	Y79F, S451N	M36I, H69K, L89M K20I, V82I	ninguna	T124A, VY201I, T206S	2	3	2	0	3
CRF15_01B	G62E, Y79F, T81A, S373A, A374N, I437L, L449P	M36I, H69K, L89M	NA	I72V, T124A	7	3	0	NA	2
CRF16_A2D	NA	NA	ninguna	L101I, T124A	NA	NA	NA	0	2
CRF17_BF	NA	NA	NA	L101I, VY201I	NA	NA	NA	NA	2
CRF18_cpx	NA	NA	NA	T124A	NA	NA	NA	NA	1

Variante del VIH-1	Polimorfismos naturales asociados a MRD (en ≥90% de secuencias)				Nº MRD-IP en Gag	Nº MRD-IP en PR		Nº MRD -RT	Nº MRD -INI
	En Gag *	En PR**	En RT***	En IN		secundarias	otras		
CRF19_cpx	NA	NA	NA	I72V, T124A, VY201I, T206S	NA	NA	NA	NA	4
CRF20_BG	NA	NA	NA	T124A, VY201I, T206S	NA	NA	NA	NA	3
CRF21_A2D	NA	NA	ninguna	I72V, VY201I	NA	NA	NA	0	2
CRF22_01A1	NA	NA	ninguna	I72V, T124A, VY201I	NA	NA	NA	0	3
CRF23_BG	NA	NA	NA	L101I, K156N	NA	NA	NA	NA	2
CRF24_BG	NA	NA	NA	L101I, T124A, VY201I, T206S	NA	NA	NA	NA	4
CRF25_cpx	G62E, R76K, S373A, S451N	M36I, I64M, H69K, L89M K20I, V82I	ninguna	T124A, VY201I, T206S	4	4	2	0	3
CRF26_AU	R76K, S373NA, L449P, P453LI	G16E, M36I, I62V, L89M, K20R	ninguna	I72V, L101I, T124A	4	4	1	0	3
CRF27_cpx	NA	NA	ninguna	VY201I, T206S	NA	NA	NA	0	2
CRF28_BF	S373AT	M36I	ninguna	ninguna	1	1	0	0	0
CRF29_BF	R76K, S373AT, S451N	M36I, L89M	ninguna	L101I	3	2	0	0	1
CRF31_BC	NA	NA	NA	L101I, VY201I	NA	NA	NA	NA	2
CRF32_06A1	NA	NA	NA	L74AI, L101I, T124A, VY201I, T206S	NA	NA	NA	NA	5
CRF33_01B	S373TA, A374G, I437L	M36I, H69K, L89M	ninguna	I72V, VY201I	3	3	0	0	2
CRF34_01B	NA	NA	NA	I72V, T124A, VY201I	NA	NA	NA	NA	3

Variante del VIH-1	Polimorfismos naturales asociados a MRD (en ≥90% de secuencias)				Nº MRD-IP en Gag	Nº MRD-IP en PR		Nº MRD -RT	Nº MRD -INI
	En Gag*	En PR**	En RT***	En IN		secundarias	otras		
CRF35_AD	G62E, Y79F, S373TA, L449P	M36I, H69K, L89M K20R	V179I	I72V, VY201I	4	3	1	1	2
CRF36_cpx	NA	NA	ninguna	L74AI, VY201I, T206S	NA	NA	NA	0	3
CRF37_cpx	NA	NA	ninguna	I72V, T124A, VY201I, T206S	NA	NA	NA	0	4
CRF38_BF	NA	NA	NA	I72V, L101I, T124A, VY201I	NA	NA	NA	NA	4
CRF39_BF	NA	NA	NA	I72V	NA	NA	NA	NA	1
CRF40_BF	NA	NA	NA	I72V, VY201I	NA	NA	NA	NA	2
CRF42_BF	NA	NA	NA	L101I, VY201I	NA	NA	NA	NA	2
CRF43_02G	G62E, R76K, Y79F, V128del, S451N	M36I, H69K, L89M K20I, V82I	ninguna	I72V, L101I, T124A, VY201I, T206S	5	3	2	0	5
CRF44_BF	NA	NA	NA	G163RK, I72V, T124A, VY201I	NA	NA	NA	NA	4
CRF45_cpx	NA	NA	NA	T124A, VY201I	NA	NA	NA	NA	2
CRF46_BF	NA	NA	NA	I72V, L101I, T124A, VY201I	NA	NA	NA	NA	4
CRF47_BF	A374N, S451N	M36I, H69K	ninguna	M154I , V165I	2	2	0	0	2
CRF49_cpx	G62E, V128I, S451N	M36I, H69K, L89M, I93L	ninguna	L101I, T124A, VY201I	3	4	0	0	3
CRF50_A1D	NA	NA	NA	VY201I	NA	NA	NA	NA	1
CRF52_01B	G62E, R76K, L449P	M36I, H69K, L89M K20R	K238R	I72V, L101I, VY201I	3	3	1	1	3
CRF53_01B	G62E, R76K, T79F, S37 3T, A374G, I437L	G16E, M36I, H69K, L89M	K238R	I72V, T124A, VY201I	6	4	0	1	3

Variante del VIH-1	Polimorfismos naturales asociados a MRD (en ≥90% de secuencias)				Nº MRD-IP en Gag	Nº MRD-IP en PR		Nº MRD -RT	Nº MRD -INI
	En Gag*	En PR**	En RT***	En IN		secundarias	otras		
CRF55_01B	NA	NA	V179E	L101I,T124A, VY201I	NA	NA	NA	1	3
CRF56_cpx	G62E,R76K,Y79F, M378I,L449P	M36I, I64V, H69K, L89M K20I	ninguna	I72V, L101I,T124A, VY201I,T206S	5	4	1	0	5
CRF63_02A1	G62E,R76K,Y79F, V128del,A374P, K436R,I437L,L449P	G16E, M36I, I64M, H69K, L89M K20I	V179E K238R	VY201I	8	5	1	2	1
CRF71_BF1	NA	NA	ninguna	L101I,V165I	NA	NA	NA	0	2

NA, polimorfismos no analizados en la región indicada. Ninguna, ninguna MRD encontrada como polimorfismo natural en ausencia de TAR en más del 90% de las secuencias de esa región viral de la variante del VIH-1 concreta. MRD-IP secundarias o raras en Gag y PR descritas en Figura 2.1A y Tabla 2.3, respectivamente. MRD en RT a ITIAN e ITINAN mostradas en Figura 3.1. Cuando el cambio incluye 2 letras se refiere a que la suma de las secuencias con esas MRD correspondía a ≥90% de sus secuencias, poniendo el primer lugar el aa más frecuente. Indicadas en azul, las MRD consideradas accesorias o secundarias. En verde se muestran las variantes del VIH-1 que no estaban analizadas en las 4 regiones virales (Gag, PR, RT, IN). Sombreados en amarillo se indican aquellas MRD encontradas sólo en una variante del VIH-1: K156N en IN de CRF23_BG; M154I en IN de CRF47_BF, E157Q en IN de CRF03_AB y V165I en IN de grupo N.

* MRD-IP en Gag ubicados dentro de un *cleavage site* mostrados en la Fig 1.3, excepto aquellos en residuos 62, 76, 79 y 81 que se localizan en P17 y fuera de los CS.

** MRD-IP en PR secundarias (en negro) u “otras” o “raras” (en azul) según clasificación de Stanford.

***MRD-RT accesorias a ITIAN (inhibidores de la RT análogos de nucleósido) o ITINAN (inhibidores de la RT no análogos de nucleósido).

La visión conjunta de todos los resultados obtenidos en esta Tesis que nos proporciona la Tabla D.1. nos permite una discusión mas con bibliografía ya existente. Así, las MRD-IP A71T o T74S estabilizan a L89I en la proteasa del subtipo G. La asociación de 89I con 90M puede aumentar aún más la resistencia a IP de virus de subtipo G cuando se compara con 90M solo, resaltando nuevos perfiles mutacionales para resistencia a fármacos en este subtipo no-B (González *et al.*, 2008). M89I/V se asoció fuertemente con las MRD-IP en los codones 71, 74 y 90. Fenotípicamente, 89I/V por sí solo no confiere una susceptibilidad reducida a IP, pero combinado con 90M redujo significativamente su susceptibilidad a NFV ($p < 0,05$) en comparación con cepas con 90M solo (Abecasis *et al.*, 2005). Estas MRD han sido encontradas también en nuestras secuencias de pacientes *naïve* del GenBank, A71T en el recombinante CRF07_BC (5,6%) y en los subtipos B (2,32%), C (0,36%); y T74S en el recombinante CRF29_BF (25%) y en los subtipos C (10,3%), B (0,1%). Por otro lado, M89I apareció de manera natural en los grupos O (100%) y P (100%), subtipo C (1,8%), subtipo B (0,12%), CRF01_AE (0,87%) (**Figura 2.1**).

Otro estudio mostró que las mutaciones M36I y A71V proporcionaban un mayor nivel de resistencia cruzada a varios IP cuando aparecían combinadas con la mutación D30N en el sitio activo de la PR, permitiendo escapar de la unión del IP pero manteniendo la eficiencia catalítica de la PR (Clemente *et al.*, 2003). Por ello, la presencia de M36I y A71V como polimorfismos naturales que hemos visto en ciertas variantes del VIH-1 como el grupo O (**Fig.2.1, y Tabla D.1**), podrían sugerir una menor susceptibilidad a IP en estas cepas. Otros trabajos revelan que la doble mutación de PR V77I+L33F ocasiona una menor unión al IP NFV y la triple mutación PR V77I+K20T+L33F una mayor unión al NFV (Gupta *et al.*, 2015). Sin embargo, este IP ya no se emplea en la práctica clínica ya que la EMA suspendió la autorización de comercialización en el año 2007. V77I fue hallada en nuestras secuencias de pacientes *naïve* del GenBank en los subtipos D (100%), B (20,7%), C (4,6%) y en las variantes recombinantes CRF07BC (38,9%), CRF29BF (25%), CRF01AE (3,4%); y L33F en CRF01AE (4,1%) y subtipo B (0,5%) (**Figura 2.1**).

Además, puesto que las mutaciones dentro de los CS han sido asociadas también con la exposición al tratamiento con IP como hemos explicado en el Capítulo 2 de esta Tesis y a la resistencia a inhibidores de la maduración (Fun *et al.*, 2012; Malet *et al.*,

2007), nuestros resultados podrían potencialmente proporcionar un mejor entendimiento del papel de *gag* en las resistencias a ARV y en el desarrollo de futuros inhibidores de la maduración (Waheed *et al.*, 2012).

Sin embargo, los polimorfismos pueden afectar la efectividad del TAR de forma diferente en grupos de pacientes específicos, como se ha visto los pacientes experimentados a ARV y/o pacientes tratados con IP distintos a LPV (Champenois *et al.*, 2008). También el VIH-2 contiene numerosos polimorfismos naturales en su PR y RT implicados en la resistencia a ciertos IP e ITINAN (Ntemgwa *et al.*, 2007; Ntemgwa *et al.*, 2009; Smith *et al.*, 2015).

La resistencia a IP podría verse aumentada en determinadas variantes del VIH-1 debido a diferencias en sustituciones secundarias de la PR una vez son seleccionadas sustituciones principales relevantes, como mostró un estudio que comparó la barrera genética entre subtipos empleando unas 2000 secuencias del VIH-1 (incluyendo más de 600 secuencias de subtipos no-B) aisladas de pacientes *naïve* a ARV en Europa (van de Vijver *et al.*, 2006).

Los polimorfismos en PR también se han encontrado en variantes diferentes al grupo mayoritario M del VIH-1. Así, un trabajo (Luk *et al.*, 2001) que analizó los genes de PR de 28 pacientes infectados por el grupo O del VIH-1 *naïve* a IP identificó en la mayoría de las cepas los residuos 10I, 15V, 36I, 41K, 62V, 63T/A/K/I, 64V, 71V y 93L. Aunque nosotros únicamente analizamos dos PR del grupo O también identificamos en ellas los polimorfismos 10I, 36I, 62V, 64V, 71V, 93L.

Esta Tesis también aporta datos sobre las MRD para las 4 familias de ARV estudiadas (IP, ITIAN, ITINAN e INI) que no aparecieron en ninguna de las secuencias de pacientes *naïve* descargadas del GenBank analizadas para esta Tesis, resumidos en la **Tabla D.2**. Por tanto, estas MRD serían exclusivamente seleccionadas por presión farmacológica y no por variabilidad natural, sugiriendo que su presencia podría disminuir la eficacia biológica del virus que la lleva o tener un efecto deletéreo significativo sobre la función de la proteína específica (Dunn *et al.* 2015) y, por tanto, sobre el ciclo viral (**Tabla D.2**). Por ejemplo, las deleciones e inserciones en regiones que codifican las proteínas

virales, particularmente deleciones e inserciones que causan cambios del marco de lectura, son relativamente raras, como también confirman nuestros resultados.

Tabla D.2. MRD que no aparecen en ninguna de las secuencias de pacientes *naïve* analizadas para esta Tesis

Secuencias analizadas (GenBank)	Nº pacientes	Nº variantes* VIH-1 diferentes	Proteína VIH-1	Tipo	MRD totalmente ausentes en pacientes <i>naïve</i>
2.141	741	28	Gag	MRD-IP	ninguna
2.141	741	28	PR	MRD-IP	L76V (primaria) K20T, L23I, E34Q y T74P (secundarias)
3.592	3.592	40	RT	MRD-ITIAN	K70E y Y115F (principales) E40F, M41I y K219W (accesorias) K65E, K65N y Q151L (raras) Deleciones en residuos 67, 68, 69, 70 de RT Inserción en residuos 69 de RT
3.592	3.592	40	RT	MRD-ITINAN	K103H/T, E138R, Y181F/S/G, G190C/T/V/E/Q, F227C (raras)
6.706	3.791	67	IN	MRD-INI	G118R, F121Y, P145S y Q146P (raras)

*Grupos, subtipos y recombinantes del VIH-1.

Son necesarios más estudios acerca de la importancia biológica de la variabilidad VIH-1 asociada a variante que se encuentra en cada sitio de procesamiento de los precursores Gag y GagPol y de sus implicaciones funcionales. La influencia de cada polimorfismo natural encontrado en posiciones asociadas a resistencia o a disminución de la susceptibilidad a fármacos ARV, en la selección de otras MRD principales o accesorias y en la aparición de nuevas vías de resistencias en las variantes donde aparecen requiere más investigación. Ello incluye el análisis de un mayor nº de secuencias de algunas variantes del VIH-1 específicas representadas por pocas secuencias en nuestro estudio.

Nuestros resultados también refuerzan la necesidad de tener en cuenta la variante del VIH y sus polimorfismos en posiciones de resistencia durante el seguimiento clínico de las personas infectadas en tratamiento. Ello requeriría siempre una buena caracterización

de la variante del VIH por análisis filogenético de las secuencias virales y la inclusión de variantes no-B del VIH-1 en futuras investigaciones de aspectos básicos y clínicos de la infección por VIH-1, así como la inclusión de pacientes infectados por variantes no-B del VIH-1 en los ensayos clínicos.

Figura D.2. Polimorfismos naturales en Gag, RT, PR e IN asociados a MRD en gag y pol presentes de manera natural en $\geq 90\%$ de secuencias de las distintas variantes del VIH-1

VARIANTE VIH-1	<i>gag</i>																
	p17 (MA)					p2		p24	p7 (NC)			p1		p6			
	G62 A/S/E	R76 K	Y79 F/H	T81 A/L	V128 I/T/A/del /G/S/D	S373 A/T/P/N /G/V/I	A374 N/T/S/V/ P/H/G		M378 Q/I/V	E428 D/K/G	A431 V	K436 E/R	I437 T/V/L/F	L449 F/P/V	S451 N/G/R/K	R452 S/K	P453 A/L/T/I
GRUPOS																	
N					I	AT	T			G				P			T
O					TA		T			G				V	RK		A
P	E									G				V	K		
SUBTIPOS																	
A																	
A1	E					A								P			
A2																	
B																K	
C																	
D							VG										
F																	
F1						TA											
F2																	
G																	
H																	
J																	
K																	

VARIANTE VIH-1	<i>gag</i>																
	p17 (MA)					p2		p24 (CA)	p7 (NC)			p1		p6			
	G62 A/S/E	R76 K	Y79 F/H	T81 A/L	V128 I/T/A/del /G/S/D	S373 A/T/P/N /G/V/I	A374 N/T/S/V /P/H/G		M378 Q/I/V	E428 D/K/G	A431 V	K436 E/R	I437 T/V/L/F	L449 F/P/V	S451 N/G/R/K	R452 S/K	P453 A/L/T/I
CRF																	
01_AE	EA					ATPGVI								P			
02_AG					del									P			
03_AB																	
04_cpx																	
05_DF																	
06_cpx																	
07_BC																	
08_BC			F										F				
09_cpx																	
10_CD																	
11_cpx																	
F12_BF																	
13_cpx																	
14_BG			F												N		
15_01B	E		F	A		A	N						L	P			
16_A2D																	
17_BF																	
18_cpx																	
19_cpx																	
20_BG																	
21_A2D																	
22_01A1																	
23_BG																	

VARIANTE VIH-1	<i>gag</i>																
	p17 (MA)					p2		p24 (CA)	p7 (NC)			p1		p6			
	G62 A/S/E	R76 K	Y79 F/H	T81 A/L	V128 I/T/A/del /G/S/D	S373 A/T/P/N /G/V/I	A374 N/T/S/V /P/H/G		M378 Q/I/V	E428 D/K/G	A431 V	K436 E/R	I437 T/V/L/F	L449 F/P/V	S451 N/G/R/K	R452 S/K	P453 A/L/T/I
CRF																	
24_BG																	
25_cpx	E					A								N			
26_AU						NA								P			LI
27_cpx																	
28_BF						AT											
29_BF						AT									N		
31_BC																	
32_06A1																	
33_01B						TA	G						L				
34_01B																	
35_AD	E		F			TA								P			
36_cpx																	
37_cpx																	
38_BF																	
39_BF																	
40_BF																	
42_BF																	
43_02G	E		F		del										N		
44_BF																	
45_cpx																	
46_BF																	
47_BF							N								N		

VARIANTE VIH-1	gag																
	p17 (MA)					p2		p24	p7 (NC)			p1		p6			
	G62 A/S/E	R76 K	Y79 F/H	T81 A/L	V128 I/T/A/del /G/S/D	S373 A/T/P/N /G/V/I	A374 N/T/S/V/ P/H/G		M378 Q/I/V	E428 D/K/G	A431 V	K436 E/R	I437 T/V/L/F	L449 F/P/V	S451 N/G/R/K	R452 S/K	P453 A/L/T/
CRF																	
49_cpx	E				I										N		
50_A1D																	
52_01B	E													P			
53_01B	E		F			T	G						L				
55_01B																	
56_cpx	E		F						I					P			
63_02A1	E		F		del		P					R	L	P			
71_BF1																	

VARIANTE VIH-1	pol																																		
	PR																	RT					IN												
	L10 IV	V11 I/L	G16 E	K20 I/MV/R	L33 F/I/V	M36 I/L/V	Q58 E	D60 E	I62 V	L63 P	I64 L/M/V	H69 K/R	A71 IL/TV	V77 I	V82 I	L89 VT/M/I	I93 L/M	T69 S/A/I/E	A98 G	V179 D/E/I	Y181 C/I/V	K238 R	I72 V	L74 AI	L101 I	T124 A	S153 AYF	M154 I	K156 N	E157 Q	G163 RK	V165 I	V/Y201 I	T206 S	
GRUPOS																																			
N	V	I		IR	I	IL						R								I															
O	I					I					V	R	V			I	L			E															
P	V	I			V	L						R	V			I		S		E															
SUBTIPOS																																			
A																																			
A1						I						K				M																			
A2																																			
B																																			
C												K					L																		
D																																			
F																																			
F1						I										M																			
F2																																			
G				I		I						K				M																			
H																																			
J																																			
K																																			

VARIANTE VIH-1	pol																																		
	PR																	RT					IN												
	L10 IV	V11 I/L	G16 E	K20 I/MV/R	L33 F/I/V	M36 I/L/V	Q58 E	D60 E	I62 V	L63 P	I64 L/M/V	H69 K/R	A71 IL/TV	V77 I	V82 I	L89 VT/M/I	I93 L/M	T69 S/A/I/E	A98 G	V179 D/E/I	Y181 C/I/V	K238 R	I72 V	L74 AI	L101 I	T124 A	S153 AYF	M154 I	K156 N	E157 Q	G163 RK	V165 I	V/Y201 I	T206 S	
CRF																																			
01_AE												K				MVI																			
02_AG																M																			
03_AB																																			
04_cpx																																			
05_DF																																			
06_cpx																																			
07_BC																	L																		
08_BC						V						K				M	L																		
09_cpx																																			
10_CD																																			
11_cpx																																			
12_BF																																			
13_cpx																																			
14_BG				I		I						K				M																			
15_01B						I						K				M																			
16_A2D																																			
17_BF																																			
18_cpx																																			
19_cpx																																			
20_BG																																			
21_A2D																																			
22_01A1																																			
23_BG																																			

VARIANTE VIH-1	pol																																		
	PR																	RT					IN												
	L10 IV	V11 I/L	G16 E	K20 I/MV/R	L33 F/I/V	M36 I/L/V	Q58 E	D60 E	I62 V	L63 P	I64 L/M/V	H69 K/R	A71 IL/TV	V77 I	V82 I	L89 VT/M/I	I93 L/M	T69 S/A/I/E	A98 G	V179 D/E/I	Y181 C/I/V	K238 R	I72 V	L74 AI	L101 I	T124 A	S153 AYF	M154 I	K156 N	E157 Q	G163 RK	V165 I	V/Y201 I	T206 S	
CRF																																			
24_BG																																			
25_cpx				I		I					M	K				M																			
26_AU				R		I										M																			
27_cpx																																			
28_BF						I																													
29_BF						I										M																			
31_BC																																			
32_06A1																																			
33_01B						I						K				M																			
34_01B																																			
35_AD				R		I						K				M				I															
36_cpx																																			
37_cpx																																			
38_BF																																			
39_BF																																			
40_BF																																			
42_BF																																			
43_02G				I		I						K				M																			
44_BF																																			
45_cpx																																			
46_BF																																			
47_BF						I						K																							

VARIANTE	pol																																		
	PR																	RT					IN												
	L10 IV	V11 I/L	G16 E	K20 I/MV/R	L33 F/I/V	M36 I/L/V	Q58 E	D60 E	I62 V	L63 P	I64 L/M/V	H69 K/R	A71 IL/TV	V77 I	V82 I	L89 VT/M/I	I93 L/M	T69 S/A/I/E	A98 G	V179 D/E/I	Y181 C/I/V	K238 R	I72 V	L74 AI	L101 I	T124 A	S153 AYF	M154 I	K156 N	E157 Q	G163 RK	V165 I	V/Y201 I	T206 S	
VIH-1																																			
CRF																																			
49_cpx						I						K				M	L																		
50_A1D																																			
52_01B				R		I						K				M																			
53_01B						I						K				M																			
55_01B																			E																
56_cpx				I		I					V	K				M																			
63_02A1				I		I					M	K				M				E															
71_BF1																																			

CONCLUSIONES

“Sólo puede ser feliz siempre el que sepa ser feliz con todo”

Confucio (552-479 a.DeC.)

1. En esta Tesis hemos evaluado el grado de conservación, los polimorfismos naturales presentes en más del 90% de secuencias virales de cada variante analizada en ausencia de tratamiento antirretroviral y los aminoácidos totalmente conservados en regiones del virus de la inmunodeficiencia humana tipo 1 con importancia biológica y terapéutica. Éstas incluían 11 sitios de procesamiento proteico en los precursores virales Gag y GagPol y los residuos de proteínas virales Gag, proteasa, retrotranscriptasa e integrasa donde aparecen las mutaciones de resistencia a las principales familias de fármacos antirretrovirales. Para ello se emplearon varios miles de secuencias virales recuperadas del GenBank de pacientes *naive* a tratamiento antirretroviral de numerosas variantes del VIH-1, incluyendo grupos, subtipos y el mayor número de recombinantes circulantes analizados hasta la fecha.
2. Empleando cerca de 13.000 secuencias virales, se ha realizado la primera descripción del grado de conservación de cada uno de los 11 sitios de procesamiento proteico en las poliproteínas precursoras Gag y GagPol en cada uno de los grupos, subtipos y formas recombinantes circulantes del VIH-1, definiéndose las secuencias consenso de cada sitios de procesamiento proteico en cada variante y los residuos altamente conservados. Observamos un nivel de conservación en los sitios de procesamiento proteico distinto según las etapas de procesamiento y que parece existir poca influencia de la exposición a inhibidores de la proteasa viral en el grado de conservación de algunos sitios de procesamiento proteico.
3. Estos resultados proporcionan nuevos hallazgos que pueden ayudar a una mejor comprensión del procesamiento viral esencial para la infectividad del VIH-1. Además, podrían ayudar al desarrollo de nuevos fármacos ARV que actúan como inhibidores de la maduración y de vacunas terapéuticas para controlar el virus sin TAR dirigidas a regiones de los sitios de procesamiento proteico conservadas entre variantes.
4. Tras analizar las posiciones en las proteínas virales Gag, PR, RT e IN previamente asociadas a mutaciones de resistencia a los fármacos antirretrovirales en uso terapéutico en varios miles de secuencias del VIH recuperadas de pacientes *naive* y depositadas en el GenBank, hemos identificado aquellos residuos que están absolutamente conservados entre variantes virales, y los cambios de aa en esas

posiciones, reflejando su frecuencia de aparición en las secuencias de cada una de las variantes del VIH-1 analizados. Asimismo, identificamos los residuos asociados a mutaciones de resistencia que podrían considerarse polimorfismos naturales asociadas a grupo, subtipo o CRF al estar presentes en el 90% o más de las secuencias analizadas de cada variante de estudio en ausencia de terapia antirretroviral.

5. Encontramos que la presencia de mutaciones de resistencia primarias a inhibidores de la proteasa, inhibidores de la transcriptasa inversa análogos de nucleótidos, inhibidores de la transcriptasa inversa no análogos de nucleósidos e inhibidores de la integrasa fue rara en pacientes *naïve*, estando presentes sólo en un bajo porcentaje de secuencias de algunas variantes, excepto el cambio asociado a resistencia a inhibidores de la retrotranscriptasa no análogos de nucleósido Y181C, que aparecía en todas las variantes del grupo O analizadas. Sin embargo, demostramos que un gran número de aa considerados previamente como mutaciones de resistencia secundarias, accesorias o raras a las distintas familias de antirretrovirales son, en realidad, polimorfismos naturales de grupos, subtipos o formas recombinantes circulantes del VIH-1 concretos, presentes incluso en ausencia de tratamiento antirretroviral. El número de polimorfismos naturales asociados a mutaciones de resistencia difirió entre variantes, pudiendo considerarse marcadores genéticos de las variantes que los presentan.
6. También identificamos todas las mutaciones de resistencia en Gag, proteasa, retrotranscriptasa e integrasa ausentes en todos los virus analizados de pacientes *naïve* y que, por tanto, sólo se seleccionarían en presencia de tratamiento antirretroviral. Ello sugiere que son residuos esenciales para mantener una correcta funcionalidad de dichas proteínas virales.
7. Se requiere más investigación sobre el efecto de cada polimorfismo natural encontrado en posiciones asociadas a resistencia a fármacos antirretrovirales en la susceptibilidad a estos fármacos, en la selección de otras mutaciones de resistencia principales o accesorias, y en la aparición de nuevas vías de resistencias en las variantes. Sería necesario analizar un mayor número de secuencias de algunas

variantes del VIH-1 específicas representadas por pocas secuencias en nuestro estudio.

8. El conocimiento de los residuos conservados entre variantes del VIH-1 en los sitios de procesamiento, y en posiciones de proteínas Gag y Pol asociadas a resistencia a antirretrovirales tienen un alto interés para el diseño de cebadores y sondas para la detección y cuantificación del VIH-1, para el desarrollo de nuevos fármacos antirretrovirales y para la identificación de epítomos conservados para el desarrollo de vacunas terapéuticas. También serán necesarios más estudios acerca de la importancia biológica de la variabilidad VIH-1 asociada a variante que se encuentra en cada sitio de procesamiento viral en el ciclo viral.
9. Nuestros resultados refuerzan la necesidad de tener en cuenta la variante del VIH y sus polimorfismos en posiciones de resistencia durante el seguimiento clínico de los pacientes en terapia antirretroviral. Ello requeriría siempre una buena caracterización de la variante del VIH por análisis filogenético de las secuencias virales y la inclusión de variantes no-B del VIH-1 en futuras investigaciones de aspectos básicos y clínicos de la infección por VIH-1, así como la inclusión de pacientes infectados por variantes no-B (incluidos recombinantes) del VIH-1 en los ensayos clínicos.

BIBLIOGRAFÍA

“Es intentando lo imposible como se realiza lo posible”

Henri Barbusse (1873-1935)

1. Abecasis AB, Deforche K, Snoeck J, Bacheler LT, McKenna P, Carvalho AP et al. Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G. *AIDS* 2005; 19(16):1799-806. Erratum in: *AIDS*. 2006 Aug 22; 20(13):1789.
2. Abecasis AB, Deforche K, Snoeck J, Bacheler LT, McKenna P, Carvalho AP, Gomes P, Camacho RJ, Vandamme AM. Protease mutation M89I/V is linked to therapy failure in patients infected with the HIV-1 non-B subtypes C, F or G. *AIDS*. 2005; 19(16):1799-806. Erratum in: *AIDS*. 2006 Aug 22; 20(13):1789
3. Adamson CS, Salzwedel K and Freed EO. Virus maturation as a new HIV-1 therapeutic target. *Expert Opin Ther Targets* 2009; 13(8):895-908.
4. Adamson CS. Protease-Mediated Maturation of HIV: Inhibitors of Protease and the Maturation Process. *Mol Biol Int*. 2012; 2012:604261.
5. Aghokeng AF, Mpoudi-Ngole E, Dimodi H, Atem-Tambe A, Tongo M et al. Inaccurate diagnosis of HIV-1 group M and O is a key challenge for ongoing universal access to antiretroviral treatment and HIV 111 prevention in Cameroon. *PLoS One* 2009; 4:e7702.
6. AIDS by the numbers 2016: http://www.unaids.org/sites/default/files/media_asset/AIDS-by-the-numbers-2016_en.pdf
7. Al-Dakkak I, Patel S, McCann E, Gadkari A, Prajapati G and Maiese EM. The impact of specific HIV treatment-related adverse events on adherence to antiretroviral therapy: a systematic review and meta-analysis. *AIDS Care* 2013; 25:400-14.
8. Álvarez M, Pérez AB, Camacho-Luque R, Chueca N, Iribarren JA, Gómez-Sirvent JL et al. Evolución de las resistencias primarias y mutaciones en pacientes VIH naive en CoRIS en el periodo 2007-2015. VII Congreso Nacional de GeSIDA, 2-4 de Diciembre 2015, Madrid. Abstract PO-05.
9. Alvarez P, Martín L, Prieto L, Obiang J, Vargas A, Avedillo P et al. HIV-1 variability and viral load technique could lead to false positive HIV-1 detection and to erroneous viral quantification in infected specimens. *J Infect*. 2015; 71(3):368-76.
10. Alvarez P, Rodríguez C, Martín L, del Romero J, Holguín A. HIV-1 RNA quantification from Dried Blood Spots and plasma using the Siemens VERSANT® HIV-1 RNA 1.0 Assay (kPCR). *Journal of Virology & Antiviral Research* 2014a, 3:2. Available at: http://scitechnol.com/hiv-rna-quantification-from-dried-blood-spots-and-plasma-using-the-siemens-versant-hiv-rna-assay kj3F.php?article_id=2167

11. Asboe D, Aitken C, Boffito M, Booth C, Cane P, Fakoya A et al. British HIV Association guidelines for the routine investigation and monitoring of adult HIV-1-infected individuals 2011. *HIV Med* 2012; 13:1-44.
12. Baeten JM, Chohan B, Lavreys L, Chohan V, McClelland RS et al. HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J. Infect. Dis.* 2007; 195:1177-80.
13. Bally F, Martinez R, Peters S, Sudre P and Telenti A. Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. *AIDS Res Hum Retroviruses* 2000; 16: 1209–1213.
14. Bangsberg DR, Kroetz DL and Deeks SG. Adherence-resistance relationships to combination HIV antiretroviral therapy. *Current HIV/AIDS reports* 2007; 4:65-72.
15. Banke S, Lillemark MR, Gerstoft J, Obel N, Jørgensen LB. Positive selection pressure introduces secondary mutations at Gag cleavage sites in human immunodeficiency virus type 1 harboring major protease resistance mutations. *J Virol.* 2009; 83(17):8916-24.
16. Banke S, Lillemark MR, Gerstoft J, Obel N and Jørgensen LB. Positive selection pressure introduces secondary mutations at Gag cleavage sites in human immunodeficiency virus type 1 harboring major protease resistance mutations. *J Virol.* 2009 Sep; 83(17):8916-24.
17. Barouch DH. Challenges in the development of an HIV-1 vaccine. *Nature* 2008; 455:613-9.
18. Barré-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J et al. Isolation of a Tlymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 1983; 220:868-71.
19. Bennett DE, Camacho RJ, Otelea D, Kuritzkes DR, Hervé Fleury, Kiuchi M et al. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One* 2009; 4:e4724.
20. Blair WS, Cao J, Fok-Seang J, Griffin P, Isaacson J, Jackson RL et al. New small-molecule inhibitor class targeting human immunodeficiency virus type 1 virion maturation. *Antimicrob Agents Chemother.* 2009; 53(12):5080-7.
21. Bonner K, Mezocho A, Roberts T, Ford N, Cohn J. Viral load monitoring as a tool to reinforce adherence: a systematic review. *J Acquir Immune Defic Syndr* 2013; 64:74-8.
22. Brennan CA, Yamaguchi J, Devare SG, Foster GA and Stramer SL. Expanded evaluation of blood donors in the United States for human immunodeficiency virus type 1 non-B

- subtypes and antiretroviral drug-resistant strains: 2005 through 2007. *Transfusion* 2010; 50:2707-12.
23. Brenner B, Turner D, Oliveira M, Moisi D, Detorio M et al. A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. *AIDS* 2003; 17:F1-5
 24. Brenner BG, Oliveira M, Doualla-Bell F, Moisi DD, Ntemgwa M et al. HIV-1 subtype C viruses rapidly develop K65R resistance to tenofovir in cell culture. *AIDS* 2006; 20:F9-13.
 25. Briones C y Domingo E. Minority report: hidden memory genomes in HIV-1 quasispecies and possible clinical implications. *AIDS Rev* 2008; 10:93-109.
 26. Bruzzone B, Ventura A, Bisio F, Mbougou FA, Miguel LM, Saladini F et al. Impact of extensive HIV-1 variability on molecular diagnosis in the Congo basin. *J Clin Virol* 2010; 47(4):372-5.
 27. Carganico A, Dupke S, Ehret R et al. New dolutegravir resistance pattern identified in a patient failing antiretroviral therapy. *J Int AIDS Soc* 2014; 17: 19749.
 28. Ceccherini-Silberstein F, Van Baelen K, Armenia D, Trignetti M, Rondelez E, Fabeni L et al. Secondary integrase resistance mutations found in HIV-1 minority quasispecies in integrase therapy-naïve patients have little or no effect on susceptibility to integrase inhibitors. *Antimicrob Agents Chemother* 2010; 54(9):3938-48.
 29. Champenois K, Bocket L, Deuffic-Burban S, Cotte L, Andre P et al. Expected response to protease inhibitors of HIV-1 non-B subtype viruses according to resistance algorithms. *AIDS* 2008; 22:1087-9.
 30. Champenois K, Deuffic-Burban S, Cotte L, André P, Choisy P, Ajana F, Bocket L, Yazdanpanah Y. Natural polymorphisms in HIV-1 protease: impact on effectiveness of a first-line lopinavir-containing antiretroviral therapy regimen. *J Med Virol.* 2008 Nov; 80(11):1871-9.
 31. Cheong HT, Chow WZ, Takebe Y, Chook JB, Chan KG, Abed Al-Darraj HA et al. Genetic Characterization of a Novel HIV-1 Circulating Recombinant Form (CRF74_01B) Identified among Intravenous Drug Users in Malaysia: Recombination History and Phylogenetic Linkage with Previously Defined Recombinant Lineages. *PLoS One* 2015; 10(7):e0133883.
 32. Ciccozzi M, Santoro MM, Giovanetti M, Andrisi L, Bertoli A and Ciotti M. HIV-1 non-B subtypes in Italy: a growing trend. *New Microbiol* 2012; 35:377-86.

33. Clavel F, Guétard D, Brun-Vézinet F, Chamaret S, Rey MA, Santos-Ferreira MO et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science* 1986; 233:343-6.
34. Clemente JC, Hemrajani R, Blum LE, Goodenow MM and Dunn BM. Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N. *Biochemistry* 2003; 42(51): 15029-35.
35. Clemente JC, Hemrajani R, Blum LE, Goodenow MM, Dunn BM. Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N. *Biochemistry* 2003; 42(51):15029-35.
36. Cohen CJ, Meyers JL, Davis KL. Association between daily antiretroviral pill burden and treatment adherence, hospitalisation risk, and other healthcare utilisation and costs in a US medicaid population with HIV. *BMJ open* 2013; 3.
37. Cozzi-Lepri A, Noguera-Julian M, Di Giallonardo F, Schuurman R, Daumer M, Aitken S et al. Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: a multicohort European case-control study using centralized ultrasensitive 454 pyrosequencing. *J Antimicrob Chemother* 2015; 70:930-40.
38. Crowell TA, Gebo KA, Blankson JN, Korthuis PT, Yehia BR, Rutstein RM et al. Hospitalization Rates and Reasons Among HIV Elite Controllers and Persons With Medically Controlled HIV Infection. *J Infect Dis* 2015; 211(11):1692-702.
39. Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, Talsania S et al. Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. *Proc Natl Acad Sci USA* 2011; 108: 11530–11535.
40. Dauwe K, Mortier V, Schauliege M, Van Den Heuvel A, Fransen K, Servais JY et al. Characteristics and spread to the native population on HIV-1 non-B subtypes in two European countries with high migration rate. *BMC Infectious Diseases* 2015; 15:524.
41. de Oliveira T, Engelbrecht S, Janse van Rensburg E, Gordon M, Bishop K, Megede JZ et al. Variability at HIV-1 subtype C protease cleavage sites: an indication of viral fitness? *J Virol* 2003; 77: 9422–9430.
42. deCamp A, Hraber P, Bailer RT, Seaman MS, Ochsenbauer C et al. Global panel of HIV-1 Env reference strains for standardized assessments of vaccine-elicited neutralizing antibodies. *J Virol* 2014; 88:2489-507.

43. Delatorre E, Couto-Fernandez JC, Bello G. HIV-1 Genetic Diversity in Northeastern Brazil: High Prevalence of Non-B Subtypes. *AIDS Res Hum Retroviruses* 2017; 33(7):639-647.
44. Descamps D, Apetrei C, Collin G, Damond F, Simon F and Brun-Vézinet F. Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. *AIDS* 1998; 12:1109-11.
45. Descamps D, Delaugerre C, Masquelier B, Ruffault A, Marcelin AG et al. Repeated HIV-1 resistance genotyping external quality assessments improve virology laboratory performance. *J Med Virol.* 2006 Feb; 78(2):153-60.
46. Deshmukh L, Louis JM, Ghirlando R, Clore GM. Transient HIV-1 Gag-protease interactions revealed by paramagnetic NMR suggest origins of compensatory drug resistance mutations. *Proc Natl Acad Sci U S A.* 2016;113(44):12456-12461.
47. Díaz Pernas P, Riesco Riesco S, Larrú Martínez B, Muñoz-Fernández MA, García-Bujalance S, de José Gómez MI. False diagnosis of HIV-1. *An Pediatr (Barc)* 2006; 65(2):158-61.
48. Doherty RS, De Oliveira T, Seebregts C, Danaviah S, Gordon M and Cassol. BioAfrica's HIV-1 proteomics resource: combining protein data with bioinformatics tools. *Retrovirology* 2005; 2: 18.
49. Doyle T, Dunn DT, Ceccherini-Silberstein F, De Mendoza C, Garcia F et al. Integrase inhibitor (INI) genotypic resistance in treatment-naïve and raltegravir-experienced patients infected with diverse HIV clades. *J Antimicrob Chemother* 2015; 70:3080-6.
50. Dunn LL, Boyer PL, McWilliams MJ, Smith SJ and Hughes SH. Mutations in human immunodeficiency virus type 1 reverse transcriptase that make it sensitive to degradation by the viral protease in virions are selected against in patients. *Virology* 2015; 484:127-35.
51. Eberle J, Gürtler L. The evolution of drug resistance interpretation algorithms: ANRS, REGA and extension of resistance analysis to HIV-1 group O and HIV-2. *Intervirology* 2012; 55(2):128-33.
52. Emerman M and Malim MH. HIV-1 regulatory/accessory genes: key to unraveling viral host and host cell biology. *Science* 1998; 280:1880-4.
53. Eshleman SH, Hoover DR, Chen S, Hudelson SE, Guay LA et al. Nevirapine (NVP) resistance in women with HIV-1 subtype C, compared with subtypes A and D, after the administration of single-dose NVP. *J Infect Dis* 2005; 192: 30-6.
54. European AIDS Clinical Society (EACS) Guidelines Version 8.1 October 2016. En: http://www.eacsociety.org/files/guidelines_8.1-english.pdf

55. European Centre for Disease Prevention and Control. HIV/AIDS surveillance in Europe 2015: <http://ecdc.europa.eu/en/publications/Publications/HIV-AIDS-surveillance-Europe-2015.pdf>
56. Excler JL, Michael NL. Lessons from HIV-1 vaccine efficacy trials. *Curr Opin HIV AIDS* 2016; 11(6):607-613.
57. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* 2014; 346:56-61.
58. Fatkenheuer G, Hoffmann C, Slim J, Rouzier R, Keung A, Li J et al. Short term administration of the CCR5 antagonist vicriviroc to patients with HIV and HCV coinfection is safe and tolerable. *J Acquir Immune Defic Syndr* 2010; 53:78-85.
59. Flynn WF, Chang MW, Tan Z, Oliveira G, Yuan J, Okulicz JF et al., Deep sequencing of protease inhibitor resistant HIV patient isolates reveals patterns of correlated mutations in Gag and protease. *PLoS Comput Biol* 2015;11(4):e1004249.
60. Frange P, Galimand J, Vidal N, Goujard C, Deveau C, Souala F et al. New and old complex recombinant HIV-1 strains among patients with primary infection in 1996-2006 in France: the French ANRS CO06 primo cohort study. *Retrovirology* 2008; 5:69.
61. Fun A, van Maarseveen NM, Pokorná J, Maas RE, Schipper PJ, Konvalinka et al. HIV-1 protease inhibitor mutations affect the development of HIV-1 resistance to the maturation inhibitor bevirimat. *Retrovirology* 2011; 8:70.
62. Fun A, Wensing AM, Verheyen J and Nijhuis M. Human immunodeficiency virus Gag and protease: partners in resistance. *Retrovirology* 2012; 9: 63.
63. Gale HB, Gitterman SR, Hoffman HJ, Gordin FM, Benator DA, Labriola AM et al. Is frequent CD4+ T-lymphocyte count monitoring necessary for persons with counts ≥ 300 cells/ μ L and HIV-1 suppression? *Clin Infect Dis* 2013; 56:1340-3.
64. Gallo R, Wong-Staal F, Montagnier L, Haseltine WA and Yoshida M. HIV/HTLV gene nomenclature. *Nature* 1988; 333:504.
65. Gatell JM, Clotet B, Podzamczar D, Miró JM. Guía práctica del sida. Clínica, diagnóstico y tratamiento. 12ª Ed.2013. ISBN 9788488825100. En: <http://www.escofetzamora.com/html/cast/catalogo/doc/Guia%20Practica%20Sida%202013.pdf>

-
66. Genebat M, Ruiz-Mateos E, González-Serna A, Pulido I, Muñoz-Fernández MÁ, Ferrando-Martínez S et al. Discordance rates between Trofile test and short-term virological response to maraviroc. *Antiviral Res* 2011; 89:182-5.
 67. Geretti A. HIV-1 subtypes: epidemiology and significance for HIV management. *Curr Opin Infect Dis.* 2006; 19:1-7.
 68. Geretti AM, Harrison L, Green H, Sabin C, Hill T, et al. Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy. *Clin Infect Dis* 2009; 48:1296-305.
 69. GeSIDA (2015). Documento de prevención y tratamiento de infecciones oportunistas y otras coinfecciones en pacientes con infección por VIH. Mayo 2015. En: <http://gesida-seimc.org/wp-content/uploads/2017/02/gesida-guiasclinicas-2015-InfeccionesOportunistasyCoinfeccionesVIH.pdf>
 70. GeSIDA (2017). Documento de consenso de GeSIDA/Plan Nacional sobre el Sida respecto al tratamiento antirretroviral en adultos infectados por el virus de la inmunodeficiencia humana (Actualización enero 2017). Panel de expertos de GeSIDA y Plan Nacional sobre el Sida. En: http://www.gesida-seimc.org/guias_clinicas.php?mn_MP=406&mn_MS=407).
 71. Ghosn J, Delaugerre C, Flandre P, Galimand J, Cohen-Codar I, Raffi F, et al. Polymorphism in Gag gene cleavage sites of HIV-1 non-B subtype and virological outcome of a first-line lopinavir/ritonavir single drug regimen. *PLoS One* 2011; 6(9):e24798.
 72. Giandhari J, Basson AE, Coovadia A, Kuhn L, Abrams EJ, Strehlau R et al. Genetic Changes in HIV-1 Gag-Protease Associated with Protease Inhibitor-Based Therapy Failure in Pediatric Patients. *AIDS Res Hum Retroviruses* 2015; 31(8):776-82.
 73. Giandhari J, Basson AE, Sutherland K, Parry CM, Cane PA et al. Contribution of Gag and Protease to HIV-1 Phenotypic Drug Resistance in Pediatric Patients Failing Protease Inhibitor-Based Therapy. *Antimicrob Agents Chemother* 2016; 60(4):2248-56.
 74. Giandhari J, Basson AE, Sutherland K, Parry CM, Cane PA, Coovadia A et al. Contribution of Gag and Protease to HIV1 Phenotypic Drug Resistance in Pediatric Patients Failing Protease Inhibitor Based Therapy. *Antimicrob Agents Chemother* 2016; 60(4):2248-56.
 75. Global AIDS update 2016. Available at http://www.unaids.org/sites/default/files/media_asset/global-AIDS-update-2016_en.pdf

76. González LM, Brindeiro RM, Tarin M, Calazans A, Soares MA, Cassol S et al. In vitro hypersusceptibility of human immunodeficiency virus type 1 subtype C protease to lopinavir. *Antimicrob Agents Chemother* 2003; 47:2817-22.
77. Gonzalez LM, Santos AF, Abecasis AB, Van Laethem K, Soares EA et al. Impact of HIV-1 protease mutations A71V/T and T74S on M89I/V-mediated protease inhibitor resistance in subtype G isolates. *J Antimicrob Chemother* 2008; 61:1201-4.
78. Gonzalez LM, Santos AF, Abecasis AB, Van Laethem K, Soares EA, Deforche K, Tanuri A, Camacho R, Vandamme AM, Soares MA. Impact of HIV-1 protease mutations A71V/T and T74S on M89I/V-mediated protease inhibitor resistance in subtype G isolates. *J Antimicrob Chemother* 2008; 61(6):1201-4.
79. González-Alba JM, Holguín Á, García R, García-Bujalance S, Alonso R, Suarez A et al. Molecular surveillance of HIV-1 in Madrid, Spain: a phylogeographic analysis. *J Virol* 2011; 85:10755-63.
80. Gottesman BS, Grosman Z, Lorber M, Levi I, Shitrit P et al. Measurement of HIV RNA in patients infected by subtype C by assays optimized for subtype B results in an underestimation of the viral load. *J Med Virol* 2004; 73:167-71.
81. Göttlinger HG, Sodroski JG and Haseltine WA. Role of capsid precursor processing and myristoylation in morphogenesis and infectivity of HIV-1. *Proc Natl Acad Sci USA* 1989; 86:5781–5785.
82. Grierson J, Koelmeyer RL, Smith A and Pitts M. Adherence to antiretroviral therapy: factors independently associated with reported difficulty taking antiretroviral therapy in a national sample of HIV-positive Australians. *HIV Med* 2011; 12:562-9.
83. Grossman Z, Paxinos EE, Averbuch D, Maayan S, Parkin NT et al. Mutation D30N is not preferentially selected by human immunodeficiency virus type 1 subtype C in the development of resistance to nelfinavir. *Antimicrob Agents Chemother* 2004; 48:2159-65.
84. Gupta A, Jamal S, Goyal S, Jain R, Wahi D, Grover A. Structural studies on molecular mechanisms of Nelfinavir resistance caused by non-active site mutation V77I in HIV-1 protease. *BMC Bioinformatics* 2015; 16 Suppl 19:S10
85. Hemelaar J, Gouws E, Ghys PD, Osmanov S and WHO-UNAIDS Network for HIV isolation and characterisation. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS* 2011; 25:679-89.
86. Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. *J Infect* 2013; 66:391-400.

87. Hirsch MS, Gunthard HF, Schapiro JM, Brun-Vézinet F, Clotet B, Hammer SM et al. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Top HIV Med.* 2008; 16(3):266-85.
88. HIV/AIDS. WHO Fact sheet N°360. Updated July 2016: <http://www.who.int/mediacentre/factsheets/fs360/en/>.
89. Hofstra LM, Sánchez Rivas E, Nijhuis M, Bank LEA, Wilkinson E, Kelly K et al. High Rates of Transmission of Drug-resistant HIV in Aruba Resulting in Reduced Susceptibility to the WHO Recommended First-line Regimen in Nearly Half of Newly Diagnosed HIV-infected Patients. *Clin Infect Dis* 2017; 64(8): 1092-1097.
90. Hofstra LM, Sauvageot N, Albert J, Alexiev I, Garcia F, Struck D et al. Transmission of HIV Drug Resistance and the Predicted Effect on Current First-line Regimens in Europe. *Clin Infect Dis* 2016; 62(5):655-663.
91. Hofstra LM, Sauvageot N, Albert J, Alexiev I, Garcia F, Struck D et al.; Transmission of HIV Drug Resistance and the Predicted Effect on Current First-line Regimens in Europe. *Clin Infect Dis.* 2016; 62(5):655-63.
92. Hogg RS, Heath K, Bangsberg D, Yip B, Press N, O'Shaughnessy MV et al. Intermittent use of triple-combination therapy is predictive of mortality at baseline and after 1 year of follow-up. *AIDS* 2002; 16:1051-8.
93. Holguín A, Alvarez A y Soriano V. Differences in the length of gag proteins among different HIV type 1 subtypes. *AIDS Res Hum Retroviruses* 2005; 21:886-887.
94. Holguín Á, de Mulder M, Yebra G, López M and Soriano V. Increase of non-B subtypes and recombinants among newly diagnosed HIV-1 native Spaniards and immigrants in Spain. *Curr HIV Res* 2008a; 6:327-34.
95. Holguín Á, Lospitao E, López M, de Arellano ER, Pena MJ et al. Genetic characterization of complex interrecombinant HIV-1 strains circulating in Spain and reliability of distinct rapid subtyping tools. *J Med Virol* 2008b; 80:383-91.
96. Holguín A, Pena MJ, Troncoso F and Soriano V Introduction of non-B subtypes among Spaniards newly diagnosed with HIV-1 in the Canary Islands. *AIDS Res Hum Retroviruses* 2007; 23:448-502.
97. Holguín Á, Ramírez de Arellano E, Rivas P y Soriano V. Efficacy of antiretroviral therapy in individuals infected with HIV-1 non-B subtypes. *AIDS Rev* 2006a; 8:98-107.

98. Holguín Á, Suñé C, Hamy F, Soriano V y Klimkait T. Natural polymorphisms in the protease gene modulate the replicative capacity of non-B HIV-1 variants in the absence of drug pressure. *J Clin Virol* 2006b; 36:264-71.
99. Holguín et al., 2008c Holguín Á, López M, Molinero M y Soriano V Performance of three commercial viral load assays, Versant human immunodeficiency virus type 1 (HIV-1) RNA bDNA v3.0, Cobas AmpliPrep/Cobas TaqMan HIV-1, and NucliSens HIV-1 EasyQ v1.2, testing HIV-1 non-B subtypes and recombinant variants. *J Clin Microbiol* 2008c; 46:2918-23.
100. Hu Y, Wan Z, Zhou YH, Smith D, Zheng YT and Zhang C. Identification of two new HIV-1 circulating recombinant forms (CRF87_cpx and CRF88_BC) from Reported Unique Recombinant Forms in Asia. *AIDS Res Hum Retroviruses* 2017; 33(4):353-358.
101. Hu Y, Wan Z, Zhou YH, Smith D, Zheng YT, Zhang C. Identification of Two New HIV-1 Circulating Recombinant Forms (CRF87_cpx and CRF88_BC) from Reported Unique Recombinant Forms in Asia. *AIDS Res Hum Retroviruses* 2017; 33(4):353-358.
102. Huang W, De Gruttola V, Fischl M, Hammer S, Richman D et al. Patterns of plasma human immunodeficiency virus type 1 RNA response to antiretroviral therapy. *J Infect Dis* 2001; 183:1455-65.
103. Hughes CA, Foisy MM, Dewhurst N, Higgins N, Robinson L, Kelly DV et al. Abacavir hypersensitivity reaction: an update. *The Annals of pharmacotherapy* 2008; 42:387-96.
104. Insight Start Study Group, Lundgren JD, Babiker AG, Gordin F, Emery S, Grund B, et al. Initiation of antiretroviral therapy in early asymptomatic HIV infection. *N Engl J Med* 2015; 373:795-807.
105. Jetzt AE, Yu H, Klarmann GJ, Klarmann G, Ron Y, Preston BD et al. High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J Virol* 2000; 74:1234-40.
106. Jiamsakul A, Sungkanuarph S, Law M, Kantor R, Praparattanapan J, Li PC et al. HIV multi-drug resistance at first-line antiretroviral failure and subsequent virological response in Asia. *J Int AIDS Soc* 2014; 17:19053.
107. Jiamsakul A, Sungkanuparph S, Law M, Kantor R, Praparattanapan J, Li PC et al. HIV multi-drug resistance at first-line antiretroviral failure and subsequent virological response in Asia. *J Int AIDS Soc* 2014; 17:19053.
108. Junqueira DM and Almeida SE. HIV-1 subtype B: Traces of pandemic. *Virology*. 2016 Aug; 495:173-84.

109. Kantor R y Katzenstein D. Polymorphism in HIV-1 non-subtype B protease and reverse transcriptase and its potential impact on drug susceptibility and drug resistance evolution. *AIDS Rev* 2003; 5:25-35.
110. Kaplan A, Manchester M and Swanstrom R. The activity of the protease of HIV-1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. *J Virol* 1994; 68:6782–6786.
111. Karmochkine M, Si Mohamed A, Piketty C, Ginsburg C, Raguin G, Schneider-Fauveau V, Gutmann L, Kazatchkine MD, Belec L. The cumulative occurrence of resistance mutations in the HIV-1 protease gene is associated with failure of salvage therapy with ritonavir and saquinavir in protease inhibitor-experienced patients. *Antiviral Res.* 2000 Sep;47(3):179-88.
112. Karmochkine M, Si Mohamed A, Piketty C, Ginsburg C, Raquin G, Schneider-Fauveau V et al. The cumulative occurrence of resistance mutations in the HIV-1 protease gene is associated with failure of salvage therapy with ritonavir and saquinavir in protease inhibitor-experienced patients. *Antiviral Res.* 2000; 47(3): 179-88.
113. Kaufmann GR, Suzuki K, Cunningham P, Mukaide M, Kondo M, Imai M et al. Impact of HIV type 1 protease, reverse transcriptase, cleavage site, and p6 mutations on the virological response to quadruple therapy with saquinavir, ritonavir, and two nucleoside analogs. *AIDS Res Hum Retroviruses* 2001; 17(6):487-97.
114. Kieffer TL, Finucane MM, Nettles RE, Quinn TC, Broman KW, Ray SC et al. Genotypic analysis of HIV-1 drug resistance at the limit of detection: virus production without evolution in treated adults with undetectable HIV loads. *J Infect Dis* 2004; 189:1452-65.
115. Kimmel AD, Goldie SJ, Walensky RP, Losina E, Weinstein MC, Patiel AD et al. Optimal frequency of CD4 cell count and HIV RNA monitoring prior to initiation of antiretroviral therapy in HIV-infected patients. *Antivir Ther* 2005; 10:41-52.
116. Kinomoto M, Appiah-Opong R, Brandful JA, Yokoyama M, Nii-Trebi N, Ugly-Kwame E et al. HIV-1 proteases from drug-naïve West African patients are differentially less susceptible to protease inhibitors. *Clin Infect Dis* 2005; 41:243-51.
117. Kiwanuka N, Laeyendecker O, Quinn TC, Wawer MJ, Shepherd J et al. HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. *AIDS* 2009; 23:2479-84.

118. Knops E, Kemper I, Schülter E, Pfister H, Kaiser R and Verheyen J. The evolution of protease mutation 76V is associated with protease mutation 46I and gag mutation 431V. *AIDS* 2010; 24(5):779-81.
119. Koning FA, Castro H, Dunn D, Tilston P, Cane PA, Mbisa JL et al. Subtype-specific differences in the development of accessory mutations associated with high-level resistance to HIV-1 nucleoside reverse transcriptase inhibitors. *J Antimicrob Chemother* 2013; 68(6):1220-36.
120. Konou AA, Vidal N, Salou M, Anato S, Singo-Tokofaï A, Ekouevi DK et al. Genetic diversity and transmission networks of HIV-1 strains among men having sex with men (MSM) in Lomé, Togo. *Infect Genet Evol* 2016; 46:279-285.
121. Korber B y Gnanakaran S. The implications of patterns in HIV diversity for neutralizing antibody induction and susceptibility. *Curr Opin HIV AIDS* 2009; 4:408-17.
122. Kouri V, Khouri R, Alemán Y, Abrahantes Y, Vercauteren J, Pineda-Peña AC et al. CRF19_cpx is an Evolutionary fit HIV-1 Variant Strongly Associated With Rapid Progression to AIDS in Cuba. *EBioMedicine* 2 2015; 244–254.
123. Kozísek M, Henke S, Sasková KG, Jacobs GB, Schuch A et al. Mutations in HIV-1 gag and pol compensate for the loss of viral fitness caused by a highly mutated protease. *Antimicrob Agents Chemother* 2012; 56: 4320–4330.
124. Kumar S, Tamura K, Nei M. MEGA: molecular evolutionary genetics analysis software for microcomputers. *Comput Appl Biosci*. 1994; 10:189–191.
125. Kuritzkes DR, Lalama CM, Ribaldo HJ, Marcial M, Meyer WA 3rd, Shikuma C et al. Preexisting resistance to nonnucleoside reverse-transcriptase inhibitors predicts virologic failure of an efavirenz-based regimen in treatment-naïve HIV-1-infected subjects. *J Infect Dis* 2008; 197:867-70.
126. Lambert-Niclot S, Flandre P, Malet I, Canestri A, Soulié C, Tubiana R et al. Impact of gag mutations on selection of darunavir resistance mutations in HIV-1 protease. *J Antimicrob Chemother*. 2008; 62(5):905-8.
127. Larrouy L, Chazallon C, Landman R, Capitant C, Peytavin G, Collin G et al. Gag mutations can impact virological response to dual-boosted protease inhibitor combinations in antiretroviral-naïve HIV-infected patients. *Antimicrob Agents Chemother* 2010; 54(7):2910-9.

128. Larrouy L, Lambert-Niclot S, Charpentier C, Fourati S, Visseaux B, Soulié C et al. Positive impact of HIV-1 gag cleavage site mutations on the virological response to darunavir boosted with ritonavir. *Antimicrob Agents Chemother* 2011; 55(4):1754-7.
129. Larrouy L, Lambert-Niclot S, Charpentier C, Fourati S, Visseaux B, Soulié C et al. Positive impact of HIV-1 gag cleavage site mutations on the virological response to darunavir boosted with ritonavir. *Antimicrob Agents Chemother*. 2011; 55(4):1754-7.
130. Lee SK, Potempa M, Kolli M, Özen A, Schiffer CA and Swanstrom R. Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral protease. *J Biol Chem* 2012; 287: 13279–13290.
131. Lemonovich TL, Watkins RR, Morrison CS, Kwok C, Chipato T, Musoke R et al. Differences in Clinical Manifestations of Acute and Early HIV-1 Infection between HIV-1 Subtypes in African Women. *J Int Assoc Provid AIDS Care* 2015; 14(5):415-22.
132. Levy JA. Pathogenesis of human immunodeficiency virus infection. *Microbiol Rev* 1993; 57:183-289.
133. Li F, Zoumplis D, Matallana C, Kilgore NR, Reddick M, Yunus AS et al. Determinants of activity of the HIV-1 maturation inhibitor PA-457. *Virology* 2006; 356 (1-2):217-24.
134. Li G, Verheyen J, Rhee SY, Voet A, Vandamme AM and Theys K. Functional conservation of HIV-1 Gag: implications for rational drug design. *Retrovirology* 2013;10:126.
135. Li JZ, Paredes R, Ribaud HJ, Svarovskaia ES, Metzner KJ, Kozal et al. Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: a systematic review and pooled analysis. *JAMA* 2011; 305:1327-35.
136. Liégeois F, Reteno DG, Mouinga-Ondémé A, Sica J and Rouet F. Short communication: high natural polymorphism in the gag gene cleavage sites of non-B HIV type 1 isolates from Gabon. *AIDS Res Hum Retroviruses* 2013; 29: 1179–1182.
137. Little SJ, Holte S, Routy JP, Daar ES, Markowitz M, Collier AC et al. Antiretroviral-drug resistance among patients recently infected with HIV. *N Engl J Med* 2002; 347:385-94.
138. Liu SL, Schacker T, Musey L, Shriner D, McElrath MJ, Corey L et al. Divergent patterns of progression to AIDS after infection from the same source: HIV type 1 evolution and antiviral responses. *J Virol* 1997; 71: 4284-95.

139. Llácer Delicado T, Torrecilla E, Holguín A. Deep analysis of HIV-1 natural variability across HIV-1 variants at residues associated with integrase inhibitor (INI) resistance in INI-naïve individuals. *J Antimicrob Chemother* 2016; 71:362-366.
140. Luk KC, Kaptué L, Zekeng L, Soriano V, Gürtler L, Devare SG et al . Naturally occurring sequence polymorphisms within HIV type 1 group O protease. *AIDS Res Hum Retroviruses* 2001; 17(16): 1555-61.
141. Luk KC, Kaptué L, Zekeng L, Soriano V, Gürtler L, Devare SG, Schochetman G, Hackett J Jr. Naturally occurring sequence polymorphisms within HIV type 1 group O protease. *AIDS Res Hum Retroviruses*. 2001; 17(16):1555-61.
142. Luo M, Capina R, Daniuk C, Tuff J, Peters H, Kimani M et al. Immunogenicity of sequences around HIV-1 protease cleavage sites: Potential targets and population coverage analysis for a HIV vaccine targeting protease cleavage sites. *Vaccine* 2013; 31: 3000–3008.
143. Mackie NE, Dunn DT, Dolling D, Garvey L, Harrison L et al. The impact of HIV-1 reverse transcriptase polymorphisms on responses to first-line nonnucleoside reverse transcriptase inhibitor-based therapy in HIV-1-infected adults. *AIDS* 2013; 27:2245-53.
144. Maïga AI, Malet I, Soulie C, Derache A, Koita V, Amella B et al. Genetic barriers for integrase inhibitor drug resistance in HIV type-1 B and CRF02_AG subtypes. *Antivir Ther* 2009; 14: 123–9.
145. Malet I, Fourati S, Charpentier C, Morand-Joubert L, Armenia D, Wirten M et al. The HIV-1 integrase G118R mutation confers raltegravir resistance to the CRF02_AG HIV-1 subtype. *J Antimicrob Chemother* 2011; 66: 2827–30.
146. Malet I, Roquebert B, Dalban C, Wirten M, Amellal B, Agher R et al. Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *J Infect* 2007; 54: 367–374.
147. Margot NA, Wong P, Kulkarni R, White K, Porter D, Abram ME et al. Commonly Transmitted HIV-1 Drug Resistance Mutations in Reverse-Transcriptase and Protease in Antiretroviral Treatment-Naïve Patients and Response to Regimens Containing Tenofovir Disoproxil Fumarate or Tenofovir Alafenamide. *J Infect Dis* 2017; 215(6):920-927.
148. Martinez-Cajas JL, Pai NP, Klein MB, Wainberg MA. Differences in resistance mutations among HIV-1 non-subtype B infections: a systematic review of evidence (1996–2008). *J Int AIDS Soc* 2009; 12: 11.
149. McKinnon JE, Delgado R, Pulido F, Shao W, Arribas JR, Mellors JW. Single genome sequencing of HIV-1 gag and protease resistance mutations at virologic failure during the

- OK04 trial of simplified versus standard maintenance therapy. *Antivir Ther* 2011;16(5):725-32.
150. Menéndez-Arias L, Alvarez M. Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection. *Antiviral Res.* 2014; 102:70-86.
151. Merindol N and Berthoux L. Restriction Factors in HIV-1 Disease Progression. *Curr HIV Res.* 2015, 13(6):448-61.
152. Mills EJ, Nachega JB, Bangsberg DR, Singh S, Rachlis B, Wu P et al. Adherence to HAART: a systematic review of developed and developing nation patient-reported barriers and facilitators. *PLoS medicine* 2006; 3:e438.
153. Misbah M, Roy G, Shahid M, Nag N, Kumar S and Husain M. Comparative analysis of drug resistance mutations in the human immunodeficiency virus reverse transcriptase gene in patients who are non-responsive, responsive and naive to antiretroviral therapy. *Arch Virol* 2016; 161:1101-1113.
154. Monge S, Guillot V, Alvarez M, Chueca N, Stella N, Peña A et al. Clinically relevant transmitted drug resistance to first line antiretroviral drugs and implications for recommendations. *PLoS One.* 2014; 9(3): e9071.
155. Monge S, Guillot V, Alvarez M, Peña A, Viciano P, García-Bujalance S et al. Analysis of transmitted drug resistance in Spain in the years 2007-2010 documents a decline in mutations to the non-nucleoside drug class. *Clin Microbiol Infect* 2012; 18: E485-90.
156. Muesing MA, Smith DH, Cabradilla CD, Benton CV, Lasky LA and Capon DJ. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* 1985; 313:450-8.
157. Müller B, Patschinsky T and Kräusslich HG. The late-domain-containing protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. *J Virol* 2002; 76: 1015–1024.
158. Munir S, Thierry E, Malet I, Subra F, Calvez V, Marcelin AG et al. G118R and F121Y mutations identified in patients failing raltegravir treatment confer dolutegravir resistance. *J Antimicrob Chemotherapy* 2014; 70: 739–49.
159. Myers RE, Pillay D. Analysis of natural sequence variation and covariation in human immunodeficiency virus type 1 integrase. *J Virol* 2008; 82:9228–35.

160. Nachega JB, Parienti JJ, Uthman OA, Gross R, Dowdy DW, Sax PE et al. Lower pill burden and once-daily antiretroviral treatment regimens for HIV infection: A meta-analysis of randomized controlled trials. *Clin Infect Dis* 2014; 58:1297-307.
161. Nettles RE, Kieffer TL, Simmons RP, Cofrancesco J Jr, Moore RD, Gallant JE et al. Genotypic resistance in HIV-1-infected patients with persistently detectable low-level viremia while receiving highly active antiretroviral therapy. *Clin Infect Dis* 2004; 39:1030-7.
162. Nijhuis M, van Maarseveen NM, Lastere S, Schipper P, Coakley E et al. A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* 2007; 4(1):e36.
163. Ntemgwa M, Brenner BG, Oliveira M, Moisi D, Wainberg MA. Natural polymorphisms in the human immunodeficiency virus type 2 protease can accelerate time to development of resistance to protease inhibitors. *Antimicrob Agents Chemother* 2007; 51(2):604-10.
164. Olender SA, Taylor BS, Wong M, Wilkin T. CROI 2015: Advances in Antiretroviral Therapy. *Top Antivir Med*. 2015 Mar-Apr; 23(1):28-45.
165. On the fast track to end AIDS. 2016. UNAIDS 2016-2022 strategy. Available at http://www.unaids.org/sites/default/files/media_asset/20151027_UNAIDS_PCB37_15_18_EN_rev1.pdf
166. Palm AA, Esbjörnsson J, Månsson F, Kvist A, Isberg PE, Biague A et al. Faster progression to AIDS and AIDS-related death among seroincident individuals infected with recombinant HIV-1 A3/CRF02_AG compared with sub-subtype A3. *J Infect Dis*. 2014; 209(5):721-8.
167. Pant Pai N, Shivkumar S y Martinez-Cajas JL. Does genetic diversity of HIV-1 non-B subtypes differentially impact disease progression in treatment-naïve HIV-1 infected individuals?: a systematic review of evidence: 1996-2010. *J Acquir Immune Defic Syndr* 2012; 59:382-8.
168. Parkin N y Schapiro J. Antiretroviral drugs resistance in non-B subtype HIV-1, HIV-2 and SIV. *Antivir Ther* 2004; 9:3-12.
169. Parry CM, Kolli M, Myers RE, Cane PA, Schiffer C and Pillay D. Three residues in HIV-1 matrix contribute to protease inhibitor susceptibility and replication capacity. *Antimicrob Agents Chemother* 2011; 55(3):1106-13.
170. Patiño Galindo JA, Torres-Puente M, Gimeno C, Ortega E, Navarro D, Galindo MJ et al. Expansion of the CRF19_cpx Variants in Spain. *J Clin Virol* 2015; 69:146-9.

171. Peeters M, Aghokeng AF, Delaporte E. Genetic diversity among human immunodeficiency virus-1 non-B subtypes in viral load and drug resistance assays. *Clin Microbiol Infect* 2010; 16:1525-31.
172. Peeters M. Recombinant HIV sequences: Their role in the global epidemic. In: Kuiken C, Foley B, Hahn B, Korber B, McCutchan F, Marx P, editors. *Theoretical Biology and Biophysics Group*. Los Alamos NM: National Laboratory 2000. pp. 1-39-54 in *HIV Sequence Compendium* 2000. (<https://www.hiv.lanl.gov/content/sequence/HIV/REVIEWS/reviews.html>)
173. Perelson AS, Neumann AU, Markowitz M, Leonard JM and Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span and viral generation time. *Science* 1996; 271:1582-6.
174. Persaud D, Palumbo P, Ziemniak C, Chen J, Ray SC, Hughes M et al. Early archiving and predominance of nonnucleoside reverse transcriptase inhibitor-resistant HIV-1 among recently infected infants born in the United States. *J Infect Dis* 2007; 195:1402-10.
175. Pettit SC, Lindquist JN, Kaplan AH and Swanstrom R. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* 2005; 2: 66.
176. Plantier JC, Djemai M, Lemée V, Reggiani A, Leoz M, Burc L et al. Census and analysis of persistent false-negative results in serological diagnosis of human immunodeficiency virus type 1 group O infections. *J Clin Microbiol* 2009; 47(9):2906-11.
177. Poonpiriya V, Sungkanuparph S, Leechanachai P, Pasomsub E, Watitpun C, Chunchakan S et al. A study of seven rule-based algorithms for the interpretation of HIV-1 genotypic resistance data in Thailand. *J Virol Methods* 2008; 151(1):79-86.
178. Poveda E, Alcamí J, Paredes R, Córdoba J, Gutiérrez F, Llibre JM et al. Genotypic determination of HIV tropism - clinical and methodological recommendations to guide the therapeutic use of CCR5 antagonists. *AIDS Rev* 2010; 12:135-48.
179. Prabu-Jeyabalan M, Nalivaika E and Schiffer C. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* 2002; 10: 369–381.
180. Quashie PK, Oliviera M, Veres T, Osman N, Han YS, Hassounah S et al. Differential effects of the G118R, H51Y and E138K resistance substitutions in different subtypes of HIV integrase. *J Virol* 2015; 89: 3163–75.

181. Ragupathy V, Zhao J, Wood O, Tang S, Lee S, Nyambi P et al. Identification of new, emerging HIV-1 unique recombinant forms and drug resistant viruses circulating in Cameroon. *Virol J.* 2011; 8:185.
182. Recomendaciones GESIDA/SEFH/PNS para mejorar la adherencia al tratamiento antirretroviral. (Actualización Junio de 2008). En: <http://www.msssi.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/docs/recomendacionesAdherenciaTrtoAntirretroviral062008.pdf>
183. Rhee S-Y, Blanco JL, Jordan MR, Taylor J, Lemey P et al. Geographic and temporal trends in the molecular epidemiology and genetic mechanisms of transmitted HIV-1 drug resistance: an individual-patient- and sequence-level meta-analysis. *PLoS Med.* 2015; 12(4):e1001810.
184. Rhee SY, Jordan MR, Raizes E, Chua A, Parkin N, Kantor R et al. HIV-1 Drug Resistance Mutations: Potential Applications for Point-of-Care Genotypic Resistance Testing. *PLoS One* 2015c; 10(12):e0145772.
185. Rhee SY, Sankaran K, Varghese V, Winters MA, Hurt CB, Eron JJ et al. HIV-1 Protease, Reverse Transcriptase, and Integrase Variation. *J Virol* 2016b; 90(13):6058-70.
186. Richardson J. The anatomy and taxonomy of protein structure. *Adv Protein Chem* 1981; 34: 167–339.
187. Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK et al. HIV-1 nomenclature proposal. *Science* 2000; 288:55-6.
188. Rodríguez-Da Silva A, Miralles C, Ocampo A y Valverde D. [Prevalence study of the genetic markers associated with slow progression of human immunodeficiency virus type 1 in the Galician population (Northwest of Spain)]. *Enferm Infecc Microbiol Clin* 2015. En prensa.
189. Rojas Sánchez P, Dominguez S, Jiménez de Ory S, Prieto L, Rojo P, Mellado MJ et al. Trends in drug resistance prevalence, HIV-1 variants and clinical status in HIV-1- infected paediatric population in Madrid: 1993-2015 Analysis. *Pediatrics Infection Disease Journal* 2017. (In press).
190. Rossi AH, Rocco CA, Mangano A, Sen L, Aulicino PC. Sequence variability in p6 gag protein and gag/pol coevolution in human immunodeficiency type 1 subtype F genomes. *AIDS Res Hum Retroviruses* 2013; 29: 1056–1060.
191. Rotta I, Almeida SM. Genotypical diversity of HIV clades and central nervous system impairment. *Arq Neuropsiquiatr.* 2011; 69(6):964-72.

192. Rouet F, Chaix ML, Nerrienet E, Ngo-Giang-Huong N, Plantier JC, Burgard M et al. Impact of HIV-1 genetic diversity on plasma HIV-1 RNA Quantification: usefulness of the Agence Nationale de Recherches sur le SIDA second-generation long terminal repeatbased real-time reverse transcriptase polymerase chain reaction test. *J Acquir Immune Defic Syndr* 2007; 45:380-8.
193. Sadiq SK, Noé F and De Fabritiis G. Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc Natl Acad Sci USA* 2012; 109: 20449–20454.
194. Samikkannu T, Atluri VS, Arias AY, Rao KV, Mulet CT, Jayant RD et al. HIV-1 subtypes B and C Tat differentially impact synaptic plasticity expression and implicates HIV-associated neurocognitive disorders. *Curr HIV Res.* 2014; 12(6):397-405.
195. Samikkannu T, Rao KV, Salam AA, Atluri VS, Kaftanovskaya EM, Agudelo M et al. HIV Subtypes B and C gp120 and Methamphetamine Interaction: Dopaminergic System Implicates Differential Neuronal Toxicity. *Sci Rep.* 2015; 5:11130.
196. Santos AF, Tebit DM, Lalonde MS, Abecasis B, Ratcliff A, Camacho RJ et al. Effect of natural polymorphisms in the HIV-1 CRF02_AG protease on protease inhibitor hypersusceptibility. *Antimicrob Agents Chemother* 2012; 56:2719-25.
197. Sayana S y Khanlou H. Maraviroc: a new CCR5 antagonist. *Expert Rev Anti Infect Ther* 2009; 7:9-19.
198. Sayana S, Javanbakht M, Weinstein M, Khanlou H. Clinical impact and cost of laboratory monitoring need review even in resource-rich setting. *J Acquir Immune Defic Syndr* 2011; 56:e97-8.
199. Sharma AL, Singh TR, Devi KR, Singh LS. Molecular epidemiology of HIV-1 among the HIV infected people of Manipur, Northeastern India: Emergence of unique recombinant forms. *J Med Virol* 2017; 89(6):989-999.
200. Shehu-Xhilaga M, Kraeusslich HG, Pettit S, Swanstrom R, Lee JY, Marshall JA et al. Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J Virol* 2001; 75: 9156–9164.
201. Siliciano JD, Siliciano RF. Recent trends in HIV-1 drug resistance. *Curr Opin Virol* 2013; 3:487-94.
202. Smith RA, Raugi DN, Pan C, Sow PS, Seydi M, Mullins JI et al. In vitro activity of dolutegravir against wild-type and integrase inhibitor-resistant HIV-2. *Retrovirology* 2015; 12:10.

203. Snoeck J, Fellay J, Bartha I, Douek DC and Telenti A. Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. *Retrovirology* 2011; 8: 87.
204. Snoeck J, Kantor R, Shafer RW, Van Laethem K, Deforche K et al. Discordances between interpretation algorithms for genotypic resistance to protease and reverse transcriptase inhibitors of human immunodeficiency virus are subtype dependent. *Antimicrob Agents Chemother* 2006; 50:694-701.
205. Spearman P. HIV-1 Gag as an Antiviral Target: Development of Assembly and Maturation Inhibitors. *Curr Top Med Chem* 2016; 16(10):1154-66.
206. Ssemwanga D, Lihana RW, Ugoji C, Abimiku A, Nkengasong J, Dakum P et al. Update on HIV-1 acquired and transmitted drug resistance in Africa. *AIDS Rev.* 2015; 17(1):3-20.
207. Steckbeck JD, Craigo JK, Barnes CO and Montelaro RC. Highly conserved structural properties of the C-terminal tail of HIV-1 gp41 protein despite substantial sequence variation among diverse clades: implications for functions in viral replication. *J Biol Chem* 2011; 286: 27156–27166.
208. Sturdevant CB, Dow A, Jabara CB, Joseph SB, Schnell G, Takamune N et al. Central nervous system compartmentalization of HIV-1 subtype C variants early and late in infection in young children. *PLoS Pathog* 2012; 8(12):e1003094.
209. Sutherland KA, Goodall RL, McCormick A, Kapaata A, Lyagoba F, Kaleebu P et al. Gag-Protease Sequence Evolution Following Protease Inhibitor Monotherapy Treatment Failure in HIV-1 Viruses Circulating in East Africa. *AIDS Res Hum Retroviruses* 2015a; 31(10):1032-7.
210. Sutherland KA, Parry CM, McCormick A, Kapaata A, Lyagoba F, Kaleebu P, et al. Evidence for Reduced HIV-1 Drug Susceptibility without Emergence of Major Protease Mutations following Protease Inhibitor Monotherapy Failure in the SARA Trial. *PLoS ONE* 2015b; 10(9): e0137834.
211. Svicher V, D'Arrigo R, Alteri C, Andreoni M, Angarano G, Antinori A et al. Performance of genotypic tropism testing in clinical practice using the enhanced sensitivity version of Trofile as reference assay: results from the OSCAR Study Group. *New Microbiol* 2010; 33:195-206.
212. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011 Oct; 28(10):2731-9.

213. Tamura K, Stecher G, Peterson D, Filipski A and Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013 Dec; 30(12):2725-9.
214. Tang MW and Shafer RW. HIV-1 antiretroviral resistance: scientific principles and clinical applications. *Drugs* 2012; 72:e1-25.
215. Tanuri A, Vicente AC, Otsuki K, Ramos CA, Ferreira OC Jr, Schechter M, Janini LM, Pieniazek D, Rayfield MA. Genetic variation and susceptibilities to protease inhibitors among subtype B and F isolates in Brazil. *Antimicrob Agents Chemother* 1999 Feb; 43(2):253-8.
216. Tatarelli P, Taramasso L, Di Biagio A, Sticchi L, Nigro N et al. HIV-1 RNA quantification in CRF02_AG HIV-1 infection: too easy to make mistakes. *New Microbiol.* 2016; 39(2):150-2.
217. Tomasselli AG and Heinrikson RL. Specificity of retroviral proteases: an analysis of viral and nonviral protein substrates. *Methods Enzymol* 1994; 241: 279 301.
218. Tongo M, Burgers WA. Challenges in the design of a T cell vaccine in the context of HIV-1 diversity. *Viruses* 2014; 6(10):3968-90.
219. Torrecilla E, Llácer Delicado T and Holguín Á. New findings in cleavage sites variability across groups, subtypes and recombinants of human immunodeficiency virus type 1. *PLoS One* 2014; 9(2):e88099.
220. Turriziani O, Montagna C, Falasca F, Bucci M, Russo G et al. Short communication: analysis of the integrase gene from HIV type 1-positive patients living in a rural area of West Cameroon. *AIDS Res Hum Retroviruses* 2012; 28: 1729–33.
221. Urano E, Ablan SD, Mandt R, Pauly GT, Sigano DM, Schneider JP et al. Alkyl Amine Bevirimat Derivatives Are Potent and Broadly Active HIV-1 Maturation Inhibitors. *Antimicrob Agents Chemother* 2015; 60(1):190-7.
222. Van de Vijver DAMC, Wensing AMJ, Angarano G, Asjö B, Balotta C et al. Differences in the frequency of minor substitutions between HIV-1 subtypes and their potential impact on the genetic barrier for resistance to protease inhibitors. *Antivir Ther* 2005; 10:S145. <https://lirias.kuleuven.be/handle/123456789/340635>
223. van de Vijver DA, Wensing AM, Angarano G, Asjö B, Balotta C, Boeri E et al. The calculated genetic barrier for antiretroviral drug resistance substitutions is largely similar for different HIV-1 subtypes. *J Acquir Immune Defic Syndr.* 2006 Mar; 41(3):352-60.
224. van der Kuyl AC and Berkhout B. The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. *Retrovirology* 2012; 9: 92.

225. van Maarseveen NM, Andersson D, Lepšík M, Fun A, Schipper PJ, de Jong D, Boucher CA, Nijhuis M. Modulation of HIV-1 Gag NC/p1 cleavage efficiency affects protease inhibitor resistance and viral replicative capacity. *Retrovirology* 2012; 9:29.
226. van Maarseveen NM, Andersson D, Lepšík M, Fun A, Schipper PJ, de Jong D et al. Modulation of HIV-1 Gag NC/p1 cleavage efficiency affects protease inhibitor resistance and viral replicative capacity. *Retrovirology* 2012; 9:29.
227. Vandekerckhove L, Verhofstede C y Vogelaers D. Maraviroc: perspectives for use in antiretroviral-naive HIV-1-infected patients. *J Antimicrob Chemother* 2009; 63:1087-96.
228. Vandekerckhove LP, Wensing AM, Kaiser R, Brun-Vezinet F, Clotet B, De Luca A et al. European guidelines on the clinical management of HIV-1 tropism testing. *Lancet Infect Dis* 2011; 11:394-407.
229. Vasan A, Renjifo B, Hertzmark E, Chaplin B, Msamanga et al. Different rates of disease progression of HIV type 1 infection in Tanzania based on infecting subtype. *Clin. Infect. Dis.* 2006; 42:843-52.
230. Vercauteren J, Wensing AM, van de Vijver DA, Albert J, Balotta C, Hamouda O et al. Transmission of drug-resistant HIV-1 is stabilizing in Europe. *J Infect Dis* 2009; 200:1503-8.
231. Verheyen J, Litau E, Sing T, Däumer M, Balduin M, Oette M et al. Compensatory mutations at the HIV cleavage sites p7/p1 and p1/p6-gag in therapy-naive and therapy-experienced patients. *Antivir Ther.* 2006; 11(7):879-87.
232. Verheyen J, Verhofstede C, Knops E, Vandekerckhove L, Fun A et al. High prevalence of bevirimat resistance mutations in protease inhibitor-resistant HIV isolates. *AIDS* 2010; 24(5):669-73.
233. Vigilancia Epidemiológica del VIH y SIDA en España. Sistema de Información sobre nuevos diagnósticos de VIH y Registro Nacional de Casos de SIDA. Secretaría del Plan Nacional sobre el SIDA/ Centro Nacional De Epidemiología. Madrid. Actualización a 30 de Junio de 2016:
https://www.msssi.gob.es/ciudadanos/enfLesiones/enfTransmisibles/sida/vigilancia/InformeVIH_SIDA_2016.pdf
234. von Wyl V, Kouyos RD, Yerly S, Böni J, Shah C, Bürgusser P et al. The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. *J Infect Dis* 2011; 204:1095-103.

235. Wagner S, Kurz M, Klimkait T; Swiss HIV Cohort Study. Algorithm evolution for drug resistance prediction: comparison of systems for HIV-1 genotyping. *Antivir Ther* 2015; 20(6):661-5.
236. Waheed AA and Freed EO. HIV type 1 Gag as a target for antiviral therapy. *AIDS Res Hum Retroviruses* 2012; 28: 54–75.
237. Wares M, Mesplède T, Quashie PK, Osman N, Han Y and Wainberg MA. The M50I polymorphic substitution in association with the R263K mutation in HIV-1 subtype B integrase increases drug resistance but does not restore viral replicative fitness. *Retrovirology* 2014; 11: 7.
238. Wensing AM, Calvez V, Günthard HF, Johnson VA, Paredes R, Pillay D et al. 2017 Update of the Drug Resistance Mutations in HIV-1. *Top Antivir Med.* 2017; 24(4):132-133.
239. Wertheimer BZ, Freedberg KA, Walensky RP, Yazdanapah Y and Losina E. Therapeutic drug monitoring in HIV treatment: a literature review. *HIV Clin Trials* 2006; 7:59-69.
240. What's new in monitoring. Fact sheet: HIV treatment and care. WHO. Febrero 2016. Disponible en: <http://www.who.int/hiv/pub/arv/arv2015-monitoring-factsheet/en/>. WHO reference number: WHO/HIV/2015.42
241. Wittkop L, Gunthard HF, de Wolf F, Dunn D, Cozzi-Lepri A, de Luca A et al. Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): a European multicohort study. *Lancet Infect Dis* 2011; 11:363-71.
242. Witvrouw M, Pannecouque C, Switzer WM, Folks TM, De Clerq E and Heneine W. Susceptibility of HIV-2, SIV and SHIV to various anti-HIV1-compounds: implications for treatment and postexposure prophylaxis. *Antivir Ther* 2004; 9:57-65.
243. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Kunstman K et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature* 2008; 455:661-4.
244. Yebra G y Holguín Á. Epidemiología de la transmisión del virus de la inmunodeficiencia humana tipo 1 resistente a antirretrovirales en pacientes naive en España. *Med Clin (Barc)* 2010b; 135:561-7.
245. Yebra G y Holguín A. The maturation inhibitor bevirimat (PA-457) can be active in patients carrying HIV type-1 non-B subtypes and recombinants. *Antiviral Therapy* 2008; 13(8):1083-5.

246. Yebra G, de Mulder M, del Romero J, Rodríguez C y Holguín Á. HIV-1 non-B subtypes: high transmitted NNRTI-resistance in Spain and impaired genotypic resistance interpretation due to variability. *Antiviral Res* 2010a; 85:409-17.
247. Yebra G, de Mulder M, Martín L, Pérez-Cachafeiro S, Rodríguez C, Labarga P et al. Sensitivity of seven HIV subtyping tools differs among subtypes/recombinants in the Spanish cohort of naïve HIV-infected patients (CoRIS) *Antiviral Research* 2011; 89:19-25.
248. Yebra G, de Mulder M, Martín L, Rodríguez C, Labarga P, Viciano I et al. Most HIV type 1 non-B infections in the Spanish cohort of antiretroviral treatment-naïve HIV-infected patients (CoRIS) are due to recombinant viruses. *J Clin Microbiol* 2012; 50:407-13.
249. Yebra G, Delgado R, Pulido F, Rubio R, Galán JC, Moreno S et al. Different trends of transmitted HIV-1 drug resistance in Madrid, Spain, among risk groups in the last decade. *Arch Virol*. 2014; 159(5):1079-87.
250. Yebra G, Rivas P, Herrero MD, López M, de Mulder M, Puente S et al. Clinical differences and viral diversity between newly HIV-diagnosed African and non-African patients in Spain (2005-2007). *AIDS Res Hum Retroviruses* 2009; 25:37-44.
251. Zhang M, Foley B, Schultz AK, Macke JP, Bulla I, Stanke M et al. The role of recombination in the emergence of a complex and dynamic HIV epidemic. *Retrovirology* 2010; 7:25.
252. Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, Ron Y et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol* 2002; 76: 11273–11282.


ANEXOS

“La mejor vida no es la más duradera, sino más bien aquella que está repleta de buenas acciones”

Marie Curie (1867-1934)

ANEXO I

Estructura de los CRF DEL VIH-1

Familia CRF	CRF	Variantes VIH implicadas	Patrón mosaico subtipo
			
01A	CRF22_01A1	CRF01_AE, A1	
01B	CRF15_01B	CRF01_AE, B	
	CRF33_01B	CRF01_AE, B	
	CRF34_01B	CRF01_AE, B	
	CRF48_01B	CRF01, B	
	CRF51_01B	CRF01_AE, B	

CRF52_01B	CRF01_AE, B	
CRF53_01B	CRF01_AE, B	
CRF54_01B	CRF01_AE, B	
CRF55_01B	CRF01_AE, B	
CRF58_01B	CRF01_AE, B	
CRF59_01B	CRF01_AE, B	
CRF67_01B	CRF01, B	

	CRF68_01B	CRF01, B	
	CRF69_01B	CRF01, B	
	CRF74_01B	CRF01, B	
	CRF76_01B	CRF01, B	
02A	CRF63_02A1	CRF02, A1	
06A	CRF32_06A1	CRF06_cpx, A1	
AB	CRF03_AB	A, B	

AD	CRF16_A2D	A2, D	
	CRF21_A2D	A2, D	
	CRF35_AD	A, D	
	CRF50_A1D	A1, D	
AE	CRF01_AE	A, E	
AG	CRF02_AG	A, G	
	CRF30_0206	CRF02_AG, CRF06_cpx	

	CRF43_02G	CRF02_AG, G	
AU	CRF26_AU	A, U	
BC	CRF07_BC	B, C	
	CRF08_BC	B, C	
	CRF31_BC	B, C	
	CRF57_BC	B, C	
	CRF60_BC	B, C	

	CRF61_BC	B, C	
	CRF62_BC	B, C	
	CRF64_BC	B, C	
	CRF85_BC	B, C	
	CRF86_BC	B, C	
	CRF88_BC	B, C	
BF	CRF12_BF	B, F1	

	CRF17_BF	B, F1	<p>Genomic map of CRF17_BF. The map shows the following genes and their coordinates: gag (295-561), pol (2807-3710), vif (5954-6385), vpr (6479-6604), vpu (6604-6604), tat (6604-6604), rev (6604-6604), env (6604-6604), and nef (6604-6604). The B and F1 status is indicated by a blue bar (B) and a green bar (F).</p>
	CRF28_BF	B, F1	<p>Genomic map of CRF28_BF. The map shows the following genes and their coordinates: gag (632-1323), pol (2072-2272), vif (5463-5463), vpr (6479-6604), vpu (6604-6604), tat (6604-6604), rev (6604-6604), env (6604-6604), and nef (6604-6604). The B and F1 status is indicated by a blue bar (B) and a green bar (F).</p>
	CRF29_BF	B, F1	<p>Genomic map of CRF29_BF. The map shows the following genes and their coordinates: gag (632-1323), pol (2072-2272), vif (5463-5463), vpr (6479-6604), vpu (6604-6604), tat (6604-6604), rev (6604-6604), env (6604-6604), and nef (6604-6604). The B and F1 status is indicated by a blue bar (B) and a green bar (F).</p>
	CRF38_BF	B, F1	<p>Genomic map of CRF38_BF. The map shows the following genes and their coordinates: gag (790-2460), pol (3713-3713), vif (5463-5463), vpr (6479-6604), vpu (6604-6604), tat (6604-6604), rev (6604-6604), env (6604-6604), and nef (6604-6604). The B and F1 status is indicated by a blue bar (B) and a green bar (F).</p>
	CRF39_BF	B, F1	<p>Genomic map of CRF39_BF. The map shows the following genes and their coordinates: gag (654-2836), pol (3020-3020), vif (5463-5463), vpr (6479-6604), vpu (6604-6604), tat (6604-6604), rev (6604-6604), env (6604-6604), and nef (6604-6604). The B and F1 status is indicated by a blue bar (B) and a green bar (F).</p>
	CRF40_BF	B, F1	<p>Genomic map of CRF40_BF. The map shows the following genes and their coordinates: gag (654-1340), pol (2468-2631), vif (5037-5037), vpr (6479-6604), vpu (6604-6604), tat (6604-6604), rev (6604-6604), env (6604-6604), and nef (6604-6604). The B and F1 status is indicated by a blue bar (B) and a green bar (F).</p>
	CRF42_BF	B, F1	<p>Genomic map of CRF42_BF. The map shows the following genes and their coordinates: gag (767-1434), pol (2247-2247), vif (5037-5037), vpr (6479-6604), vpu (6604-6604), tat (6604-6604), rev (6604-6604), env (6604-6604), and nef (6604-6604). The B and F1 status is indicated by a blue bar (B) and a green bar (F).</p>

	CRF44_BF	B, F1	
	CRF46_BF	B, F1	
	CRF47_BF	B, F1	
	CRF70_BF	B, F1	
	CRF71_BF	B, F1	
	CRF72_BF	B, F1	
BG	CRF14_BG	B, G	

	CRF20_BG	B, G	
	CRF23_BG	B, G	
	CRF24_BG	B, G	
	CRF73_BG	B, G	
CD	CRF10_CD	C, D	
	CRF41_CD	C, D	
cpx	CRF04_cpx	A, G, H, K, U	

	CRF06_cpx	A, G, J, K	
	CRF09_cpx	A, G, U	
	CRF11_cpx	A, E, G, J, U	
	CRF13_cpx	CRF01, A, G, J, U	
	CRF18_cpx	A1, F, G, H, K, U	
	CRF19_cpx	A1, D, G	
	CRF25_cpx	A, G, U	

	CRF27_cpx	A, E, G, H, J, K, U	<p>Genomic map of CRF27_cpx. Genes: gag (893-2102), pol (2288-3887), vif (4899-5607), vpr (5987-6388), vpu (5987-6388), tat (8356-8660), rev (8660-9047), env (9047-9411), nef (9411-9411). Legend: G (green), H (yellow), U (grey), J (cyan), K (purple), 01 (yellow), A (red).</p>
	CRF36_cpx	CRF01, CRF02, A, G	<p>Genomic map of CRF36_cpx. Genes: gag (796-1330), pol (3387-4306), vif (5289-6084), vpr (6254-6548), vpu (6084-6254), tat (7239-8369), rev (8369-9189), env (9189-9189), nef (9189-9189). Legend: 01 (purple), 02 (yellow), A (red), G (green).</p>
	CRF37_cpx	CRF01, CRF02, A, G, U	<p>Genomic map of CRF37_cpx. Genes: gag (796-1126), pol (2513-4663), vif (5120-5772), vpr (6089-6431), vpu (5772-6089), tat (7743-8109), rev (8109-8612), env (8612-9189), nef (9189-9189). Legend: 01 (purple), A (red), G (green), U (grey), A1 (orange), 02 (yellow).</p>
	CRF45_cpx	A, K, U	<p>Genomic map of CRF45_cpx. Genes: gag (430-2398), pol (2398-3323), vif (4080-4299), vpr (6042-6343), vpu (6042-6343), tat (8227-8605), rev (8605-9146), env (9146-9146), nef (9146-9146). Legend: A (red), U (grey), K (purple).</p>
	CRF49_cpx	A1, C, J, K, U	<p>Genomic map of CRF49_cpx. Genes: gag (596-2391+5), pol (2391+5-3744+15), vif (5652+79), vpr (6042-6343), vpu (6042-6343), tat (7018+37), rev (8354+13), env (9078-9078), nef (9078-9078). Legend: U (grey), A (red), J (cyan), K (purple), C (brown).</p>
	CRF56_cpx	CRF02, B, G	<p>Genomic map of CRF56_cpx. Genes: gag (560-2703+48), pol (3204+26-3334+6), vif (5760+6), vpr (6380+14), vpu (5760+6), tat (7018+37), rev (8354+13), env (9078-9078), nef (9078-9078). Legend: 02 (yellow), B (blue), G (green).</p>
	CRF65_cpx	CRF01, B, C	<p>Genomic map of CRF65_cpx. Genes: gag (790-1184+13), pol (3781+20-4085+40), vif (5649+6-6001+7), vpr (6388+9), vpu (5649+6-6001+7), tat (7975+12), rev (8316+14-8578+8), env (9015+19-9613), nef (9613-9613). Legend: 01 (purple), C (brown), B (blue).</p>

	CRF78_cpx	CRF01, B, C	
	CRF82_cpx	CRF01, B, C	
	CRF83_cpx	CRF01, B, C	
	CRF87_cpx	CRF01, B, C	
DF	CRF05_DF	D, F	

ANEXO II

Fármacos antirretrovirales autorizados en Europa agrupados por familias

Familia	Principios Activos	Nombre del Medicamento	Fecha de autorización
ITIAN	Estavudina (d4T)	Zerit	08/05/1996
	Lamivudina (3TC)	Epivir Lamivudine Teva Pharma B.V. (*)	08/08/1996 10/12/2009
	Abacavir (ABC)	Ziagen	08/07/1999
	Emtricitabina (FTC)	Emtriva	24/10/2003
	Tenofovir disoproxil fumarato (TDF)	Viread	05/02/2002
	Tenofovir disoproxil	Tenofovir disoproxil Mylan (*)	08/12/2016
	Tenofovir disoproxil fostato	Tenofovir disoproxil Zentiva (*)	15/09/2016
2 ITIAN	Lamivudina/zidovudina (3TC/ZDV)	Combivir Lamivudine/Zidovudine Teva (*)	18/03/1998 28/02/2011
	Abacavir/lamivudina (ABC/3TC)	Kivexa	17/12/2004
	Emtricitabina/tenofovir disoproxil fumarato (FTC/TDF)	Truvada	21/02/2005
	Emtricitabina/tenofovir disoproxil fosfato	Emtricitabina/tenofovir disoproxil Zentiva (*)	09/11/2016
	Emtricitabina/tenofovir disoproxil succinato	Emtricitabina/tenofovir disoproxil Krka (*)	09/12/2016
	Emtricitabina/tenofovir disoproxil maleato	Emtricitabina/tenofovir disoproxil Mylan (*)	16/12/2016
	Emtricitabina/tenofovir alafenamida (FTC/TAF)	Descovy	21/04/2016
3 ITIAN	Abacavir/lamivudina/zidovudina (ABC/3TC/ZDV)	Trizivir	28/12/2000
ITINANN	Nevirapina (NVP)	Viramune Nevirapine Teva (*)	05/02/1998 30/11/2009
	Efavirenz (EFV)	Sustiva Stocrin Efavirenz Teva (*)	28/05/1999 28/05/1999 09/01/2012
	Etravirina (ETV)	Intelence	28/08/2008
	Rilpivirina hidrocloreuro (RPV)	Edurant	28/11/2011
IP	Indinavir sulfato (IDV)	Crixivan	04/10/1996
	Saquinavir (SQV)	Invirase	04/10/1996
	Atazanavir sulfato (ATV)	Reyataz Atazanavir Mylan (*)	02/03/2004 22/08/2016
	Fosamprenavir cálcico (FPV)	Telzir	12/07/2004
	Tipranavir (TPV)	Aptivus	25/10/2005
	Darunavir (DRV)	Prezista	12/02/2007
	Lopinavir/ritonavir (LPV/r)	Kaletra Lopinavir/Ritonavir Mylan(*)	20/03/2001 14/01/2016
IP potenciados con RTV o COBI	Darunavir/cobicistat (DRV/c)	Rezolsta	19/11/2014
	Atazanavir sulfato/cobicistat (ATV/c)	Evotaz	13/07/2015

Familia	Principios Activos	Nombre del Medicamento	Fecha de Autorización
Potenciadores farmacocinéticos (de LPV, ATV, DRV, EVG)	Ritonavir (RTV; r) Cobicistat (COBI; c)	Norvir Tybost	26/08/1996 19/09/2013
INSTI	Raltegravir (RAL)	Isentress	20/12/2007
	Dolutegravir (DTG)	Tivicay	16/01/2014
	Cabotegravir	No disponible	No autorizado aún
Inhibidores de la fusión	Enfuvirtida (ENF)	Fuzeon	27/05/2003
Inhibidores de la entrada	Maraviroc (MVC)	Celsentri	18/09/2007
	Fostemsavir trometamina	No disponible	No autorizado aún
Inhibidores de la maduración	Bevirimat	No disponible	No autorizados aún
	BMS-955176	No disponible	
2 ITIAN+ITINAN	Emtricitabina/tenofovir disoproxil fumarato/rilpivirina (FTC/TDF/RPV)	Eviplera	28/11/2011
	emtricitabine / tenofovir alafenamide/ rilpivirine hydrochloride (FTC/TAF/RPV)	Odefsey	21/06/2016
	Emtricitabina/tenofovir disoproxil fumarato/efavirenz (FTC/TDF/EFV)	Atripla	13/12/2007
2 ITIAN+INI	Abacavir sulfato/lamivudina/dolutegravir (ABC/3TC/DTG)	Triumeq	01/09/2014
	Emtricitabina/tenofovir disoproxil fumarato/elvitegravir/cobicistat (FTC/TDF/EVG/COBI)	Stribild	24/05/2013
	Emtricitabina/tenofovir alafenamida/elvitegravir/cobicistat (FTC/TAF/EVG/COBI)	Genvoya	19/11/2015

Fármacos antirretrovirales aprobados por la EMA agrupados por familias (<http://www.ema.europa.eu/ema/>).

(*) Medicamentos genéricos.

ITIAN: Inhibidores de la transcriptasa inversa análogos de nucleósidos y análogos de nucleótidos. ITINAN: Inhibidores de la transcriptasa inversa no nucleósidos. IP: Inhibidores de la proteasa. IP/r: Inhibidores de la proteasa potenciados con ritonavir. IP/c: Inhibidores de la proteasa potenciados con cobicistat. INI: Inhibidores de la integrasa.

ANEXO III

Combinaciones de ARV recomendadas para el TAR de inicio

Regímenes	Preferentes	Alternativas	Otras pautas posibles
2 ITIAN+INI			
	ABC/3TC+DTG	TDF/FTC/EVG/c	ABC/3TC+ RAL
	TFV*/FTC+DTG		
	TFV*/FTC+RAL		
	TAF/FTC/EVG/c		
2 ITIAN+ITINAN			
		TFV*/FTC+RPV	
		TFV*/FTC+EFV	
2 ITIAN+IP/p			
		TFV*/FTC+DRV/c	TFV*/FTC+ATV/c
			ABC/3TC+ DRV/c
INI+IP			
			RAL+ DRV/c

Guías clínicas GeSIDA 2017 (<http://gesida-seimc.org/category/guias-clinicas/antirretroviral-vigentes/>).

ABC: abacavir.

3TC: lamivudina.

DTG: dolutegravir.

FTC: emtricitabina.

TFV*: Tenofovir como TDF (tenofovir disoproxil fumarato) o cualquiera de sus otras sales (fosfato, maleato, succinato) o como TAF (tenofovir alafenamida fumarato).

RAL: raltegravir.

EVG/c: elvitegravir potenciado con cobicistat.

RPV: rilpivirina.

EFV: efavirenz.

DRV/c: darunavir potenciado con cobicistat.

ATV/c: atazanavir potenciado con cobicistat.

ANEXO IV

Publicaciones

Artículos derivados de los resultados de esta Tesis

1. Llácer Delicado T, Torrecilla E, Holguín A. Deep analysis of HIV-1 natural variability across HIV-1 variants at residues associated with integrase inhibitor (INI) resistance in INI-naive individuals. *J Antimicrob Chemother* 2016; 71:362-366.
2. Torrecilla E, Llácer Delicado T and Holguín Á. New findings in cleavage sites variability across groups, subtypes and recombinants of human immunodeficiency virus type 1. *PLoS One* 2014; 9(2):e88099.

Otros artículos

3. Torrecilla García E, Yebra Sanz G, Llácer-Delicado T, Rubio García R, González-García J, García García F et al. Clinical, epidemiological and treatment failure data among HIV-1 non-B-infected patients in the Spanish AIDS Research Network Cohort. *Enferm Infecc Microbiol Clin*. 2016; 34(6):353-60.

Deep analysis of HIV-1 natural variability across HIV-1 variants at residues associated with integrase inhibitor (INI) resistance in INI-naïve individuals

Teresa Llácer Delicado, Esther Torrecilla and África Holguín*

HIV-1 Molecular Epidemiology Laboratory, Microbiology and Parasitology Department, Hospital Universitario Ramón y Cajal, Madrid, Spain; Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain; CIBER en Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

*Corresponding author. HIV-1 Molecular Epidemiology Laboratory, Microbiology Department, Hospital Universitario Ramón y Cajal-IRYCIS and CIBERESP, Madrid, Spain. Fax: +34-91-3368153; E-mail: africa.holguin@salud.madrid.org

Received 27 May 2015; returned 14 July 2015; revised 16 August 2015; accepted 15 September 2015

Objectives: We evaluated variant-associated variability at positions related to resistance to the integrase (IN) inhibitors (INIs) raltegravir, elvitegravir and dolutegravir using HIV-1 IN sequences from naïve individuals retrieved from GenBank.

Methods: We evaluated the frequency of major, secondary and rare amino acid changes associated with INI resistance (INI-R) in 6706 sequences from 3791 INI-naïve individuals carrying a large panel of different HIV-1 variants retrieved from GenBank, including four groups: M (6663), O (24), N (15) and P (4). HIV-1 group M sequences included 4599 sequences from the nine group M subtypes and 2064 recombinants ascribed to 54 circulating recombinant forms (CRFs).

Results: Primary INI-R mutations were rare in INI-naïve participants and only present at a low rate in subtypes B, C and D and recombinants CRF01_AE and CRF14_BG, ranging from one to five per variant. Three secondary INI-R changes appeared with variable frequency in INI-naïve individuals carrying specific HIV-1 variants: L74M in CRF43_02G (33.3%); T97A in group P (50%), J (33.3%), CRF18_cpx (20%) and F2 (11.5%); and G163RK in CRF44_BF (100%), CRF46_BF (66.7%), CRF17_BF (28.6%), F1 (21.7%), CRF12_BF (16.7%) and CRF29_BF (12.5%). Rare mutations were absent.

Conclusions: Natural variability in INI-R positions across HIV-1 variants should be studied as they may facilitate or delay the emergence of INI-R viruses.

Introduction

HIV-1 integrase (IN) sequence availability from GenBank allows evaluation of variant-associated variability at positions related to primary and secondary resistance to the IN inhibitors (INIs) raltegravir, elvitegravir and dolutegravir, approved antiretroviral drugs targeting the HIV-1 IN and blocking incorporation of the viral DNA into the host cell DNA.¹ These inhibitors bind at the IN active site and inhibit DNA strand transfer.² Although structurally diverse, the three drugs bind to a common D64-D116-E152 motif in the IN catalytic domain (amino acids 51–212), causing it to disengage from the viral DNA.³ Despite their clinical efficacy, clinical resistance mutations have limited the promise of long-term drug prescription and increased the risk of virological failure.^{1–4}

Since nearly 90% of 35 million infections worldwide are due to HIV-1 group M non-B subtypes and recombinants, additional data are needed for a better understanding of the natural

polymorphisms at IN residues associated with INI resistance (INI-R) in each HIV-1 circulating variant. They could influence the genetic barrier and/or drive the selection of specific INI-R pathways, as they may facilitate or delay the emergence of variants fully resistant to INIs.⁴ To date, some reports have shown natural variability in subtype B⁵ and non-B subtypes.^{4,6–14} However, only a few studies have evaluated natural variability in INI-R residues among circulating recombinant forms (CRFs).^{4,7,8,11,12,14,15} In fact, natural INI-R variability was provided for only 8 (11%) of the 72 CRFs described to date (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>): CRF01_AE,^{4,7,8,11,14} CRF02_AG,^{7,8,12,15} CRF06_cpx,¹² CRF09_cpx,¹² CRF11_cpx,⁴ CRF12_BF,¹² CRF13_cpx⁸ and CRF14_BG,¹² including a lower number of INI-naïve individuals than the present analysis. The aim of the study was to evaluate the frequency of major, secondary and rare amino acid changes associated with INI-R in sequences from 3791 INI-naïve individuals carrying a large

Table 1. Resistance mutations responsible for INI susceptibility reduction and predicted scores according to Stanford

INI-R changes (Stanford)	INI (scores predicting resistance levels ^a)		
	raltegravir	elvitegravir	dolutegravir
Primary			
T66IAK	I (15)	I (60)	
	A (15)	A (60)	
	K (60)	K (60)	K (10)
E92QGV	Q (30)	Q (60)	Q (10)
	V (30)	V (60)	V (5)
	G (15)	G (60)	G (5)
Y143RCHASKG	R (60)	R (10)	
	C (60)	C (10)	
	H (60)	H (10)	
	A (60)	A (10)	
	S (60)	S (10)	
	K (60)	K (10)	
	G (60)	G (10)	
S147G		G (60)	
Q148HKR	H (60)	H (60)	H (20)
	K (60)	K (60)	K (20)
	R (60)	R (60)	R (20)
N155HTS	H (60)	H (60)	H (10)
	S (30)	S (30)	
	T (30)	T (30)	
Accessory			
H51Y	Y (15)	Y (15)	Y (10)
L74M	M (10)	M (10)	M (5)
T97A	A (15)	A (10)	A (5)
E138KA	K (15)	K (15)	K (10)
	A (15)	A (15)	A (10)
G140SAC	S (30)	S (30)	S (10)
	A (30)	A (30)	A (10)
	C (30)	C (30)	C (10)
V151LA	L (30)	L (60)	L (15)
	A (15)	A (30)	
S153YF		Y (15)	Y (10)
		F (10)	F (10)
G163RK	R (15)	R (15)	
	K (15)	K (15)	
S230R	R (15)	R (15)	
R263K	K (15)	K (30)	K (15)
Rare			
G118R	R (30)	R (30)	R (10)
F121Y	Y (30)	Y (30)	Y (5)
P145S		S (60)	
Q146P		P (60)	

Types of INI-R changes are in accordance with Stanford (<http://hivdb.stanford.edu/DR/INIResiNote.html> and http://hivdb.stanford.edu/pages/POC_INIDRMs_Summary.html; last updated on 2 March 2014 and 3 March 2015, respectively).

^aINI-R scores are as indicated in http://hivdb.stanford.edu/DR/cgi-bin/rules_scores_hivdb.cgi?class=INI. A penalty score of 5–29 predicts low-level resistance, 30–59 predicts intermediate resistance and 60 predicts high-level resistance.

number of different HIV-1 variants, including the largest panel of different CRF recombinants with available sequence analysed to date.

Methods

To identify the amino acid changes associated with INI-R in IN sequences retrieved from GenBank until January 2015 from INI-naïve individuals carrying different HIV-1 variants, we considered those substitutions included in the Stanford HIV Drug Resistance Database defined as major (T66IAK, E92QGV, Y143RCHASKG, S147G, Q148HKR and N155HTS), secondary (H51Y, L74M, T97A, E138KA, G140SAC, V151LA, S153YF, G163RK, S230R and R263K) and rare changes (G118R, F121Y, P145S and Q146P) (Table 1). In addition, we also evaluated the presence of an additional 15 changes (V72I, L74AI, Q95K, L101I, H114Y, S119R, T124A, T125K, A128T, E138T, Q148EG, V151I, E157Q, G193E and I203M) previously related to INI-R and reported by different sources (Table 2): Spanish AIDS Research Network (RIS; <http://retic-ris.net/Programa1/WP3.Resistencias.aspx>; 2013), French National Agency for AIDS Research (ANRS; <http://www.hivfrenchresistance.org/2014/tab6.html>), EMA (<http://www.ema.europa.eu/ema/>), Myers and Pillay⁷ or Carganico et al.¹⁶

We considered as INI naïve all participants with IN sequences with sampling year before 2007 (year of marketing authorization of first INI, raltegravir, and the start date of the first clinical study with an INI authorized both in Europe and the USA) and those sequences marked in GenBank as derived from only drug-naïve individuals from 2007 to January 2015. We retrieved all IN sequences from INI-naïve individuals after 2007 from the GenBank HIV sequence database (<http://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>) using two filters in the sequence search interface. Within the sequence information box, we

Table 2. Resistance mutations responsible for INI susceptibility reduction according to sources other than Stanford

Other INI-R changes	Affected INI	Information source
V72I	RAL	RIS
L74AI	RAL	Myers and Pillay ⁷
Q95K	RAL, EVG	RIS
L101I	DTG	RIS
H114Y	EVG	RIS
S119R	DTG	Carganico et al. ¹⁶
T124A	DTG	RIS
T125K	EVG	RIS
A128T	EVG	RIS
E138T	DTG	EMA
Q148EG	RAL	ANRS
V151I	RAL, EVG, DTG	RIS, EMA, Myers and Pillay ⁷
E157Q	RAL, EVG, DTG	RIS, ANRS, EMA, Myers and Pillay ⁷ , Carganico et al. ¹⁶
G193E	DTG	RIS, EMA
I203M	RAL	Myers and Pillay ⁷

RAL, raltegravir; EVG, elvitegravir; DTG, dolutegravir.

The EMA list proceeds from the summary of product characteristics of each INI, which includes resistance data provided by the results from major clinical studies. The summary of product characteristics is a document describing the properties and officially approved conditions of use of a medicine.

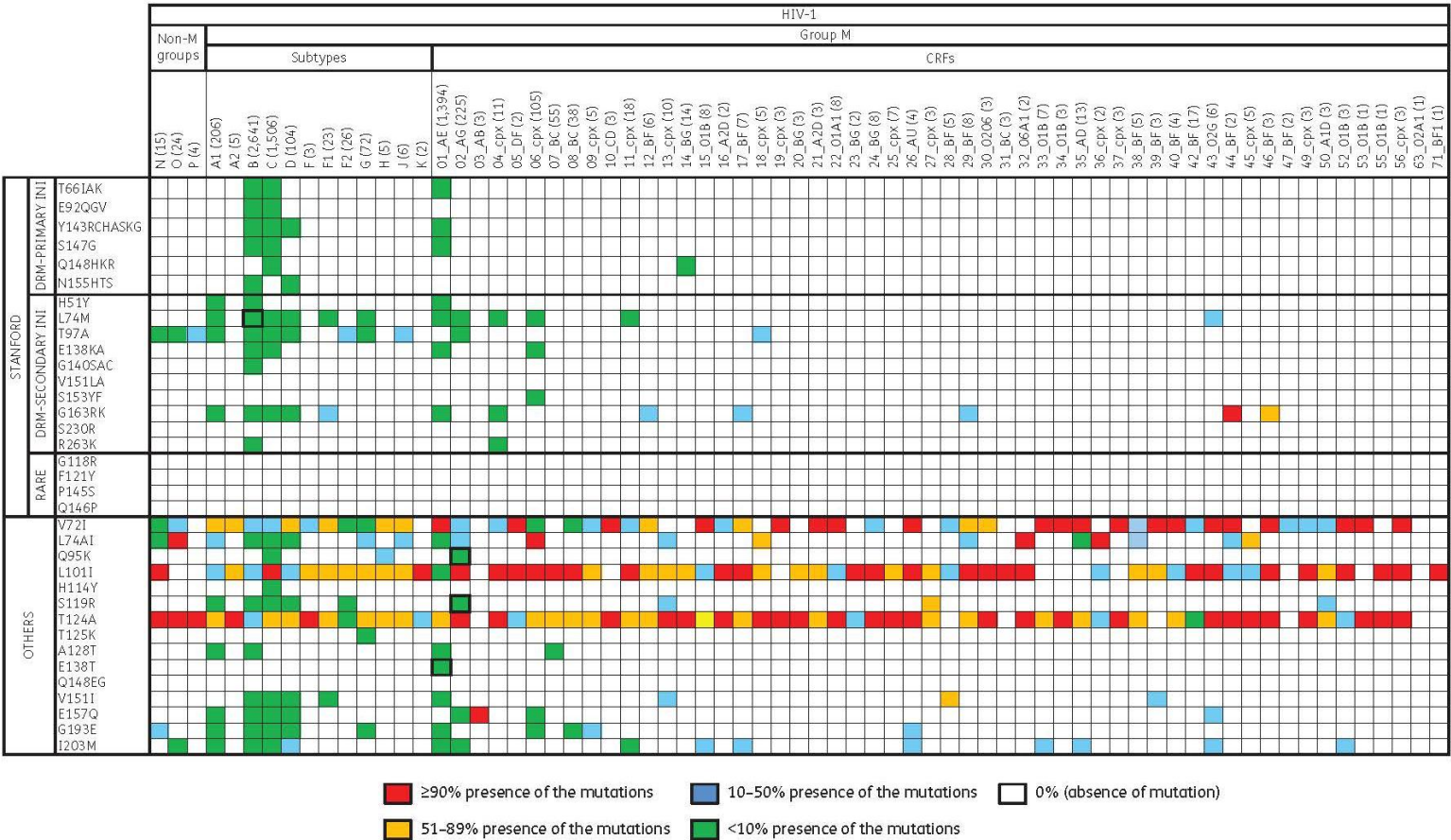


Figure 1. Natural variability in each INI-R position across HIV-1 variants. The frequency of each INI-R change in each HIV-1 variant from INI-naïve individuals was evaluated, establishing a colour code to indicate the rate of sequences with the corresponding INI-R change. In parentheses, number of IN analysed sequences. Each IN sequence was derived from individual subjects, except for: groups N (15 sequences/11 subjects), P (24 sequences/23 subjects) and O (4 sequences/3 subjects); subtypes A1 (206 sequences/177 subjects), B (2641 sequences/1131 subjects), C (1506 sequences/1010 subjects), D (104 sequences/91 subjects), F1 (23 sequences/20 subjects) and G (72 sequences/62 subjects); and CRFs: CRF01_AE (1394 sequences/590 subjects), CRF02_AG (225 sequences/205 subjects), CRF04_cpx (11 sequences/8 subjects), CRF06_cpx (105 sequences/104 subjects), CRF07_BC (55 sequences/46 subjects), CRF08_BC (38 sequences/34 subjects), CRF14_BG (14 sequences/12 subjects), CRF25_cpx (7 sequences/5 subjects), CRF33_01B (7 sequences/6 subjects) and CRF43_02G (6 sequences/4 subjects). Boxes with a bold border are changes in IN sequences but not reported after IN sequence analysis using Stanford HIVdb Program (<http://sierra2.stanford.edu/sierra/servlet/JSierra?action=sequenceInput>). DRM, drug resistance mutation.

Variability related to integrase inhibitor resistance

selected 'sampling year >=2007' and 'only drug naïve sequences'. We excluded all sequences after 2007 with no information regarding anti-retroviral treatment exposure. Thus, all retrieved IN sequences should belong to patients with no antiretroviral exposure and, thus, all would be INI-naïve individuals.

The downloaded IN nucleotide sequences were aligned and translated into amino acid sequences using MEGA version 6.0 (<http://www.megasoftware.net/>) to identify the INI-R substitutions. The frequency of each INI-R change in each HIV-1 variant from INI-naïve individuals was evaluated, establishing a colour code to indicate the rate of sequences with the corresponding INI-R change: red ($\geq 90\%$), yellow (51%–89%), blue (10%–50%), green ($<10\%$) and white (absence of mutation). Complete genomes from GenBank were used for the alignments in all HIV-1 variants, since in some of them the IN sequences were scarce or not available when only the p31-IN coding region was selected in the search.

The penalty scores for each INI-R mutation and drug were provided by Stanford (http://hivdb.stanford.edu/DR/cgi-bin/rules_scores_hivdb.cgi?class=INI). A penalty score of 5–29 predicts low-level resistance, 30–59 predicts intermediate resistance and 60 predicts high-level resistance.

Results

A total of 6706 HIV-1 IN sequences from 3791 INI-naïve individuals were retrieved from GenBank, including four groups: M (6663), O (24), N (15) and P (4). HIV-1 group M sequences included 4599 sequences from 2538 subjects from the nine group M subtypes and 2064 recombinants from 1216 INI-naïve subjects ascribed to 54 different CRFs (Figure 1). Resistance mutations responsible for raltegravir, elvitegravir and dolutegravir susceptibility reduction and penalty scores according to the Stanford database are shown in Table 1. Primary INI-R mutations were rare in INI-naïve subjects and only present at a low rate in subtypes B (0.6%), C (0.7%) and D (1.9%) and recombinants CRF01_AE (0.7%) and CRF14_BG (7.1%), ranging from one to five INI-R mutations per viral variant (Figure 1).

None of the 6706 HIV-1 IN sequences from naïve individuals carried Q148EG or secondary INI-R changes V151LA and S230R. The four rare INI-R mutations (G118R, F121Y, P145S and Q146P) were also absent across all studied HIV-1 variants (Figure 1), possibly explained by their impaired effect on the catalytic properties of IN, such as has been recently reported for G118R and F121Y.¹⁷

However, three secondary INI-R mutations (L74M, T97A and G163RK) appeared with variable frequency and $>10\%$ in INI-naïve subjects carrying specific HIV-1 variants: L74M in CRF43_Q2G (33.3%); T97A in group P (50%), J (33.3%), CRF18_cpx (20%) and sub-subtype F2 (11.5%). G163RK presented high prevalence among some BF recombinants [CRF44_BF (100%), CRF46_BF (66.7%), CRF17_BF (28.6%), sub-subtype F1 (21.7%), CRF12_BF (16.7%) and CRF29_BF (12.5%)]. The INI response in these strains carrying natural INI-R mutations might be further studied in future, including more specimens for a better understanding of their effect on INI susceptibility.

Some of the 15 INI-R mutations not included in the Stanford list but in other sources were, in fact, present in $\geq 90\%$ of sequences of specific non-B variants: V72I, L74AI, L101I, T124A and E157Q. They could be considered as natural polymorphisms in a large number of non-B group M variants (Figure 1). T124A was present in all IN sequences from groups N and P and in 96% of 24 group O sequences. L101I was highly prevalent in group N

(93.3%). L74I appeared in all group O sequences (100%) and in 6.7% of group N sequences.

Discussion

Our data reinforce that some of those reported INI-R changes not included in the Stanford database could be variant-associated polymorphisms more than INI-R mutations. However, these polymorphic substitutions should be studied further to understand their role in the evolution of INI-R and in the pathway of resistance to each INI in each HIV-1 variant.

A recent report has shown subtype-specific effects caused by polymorphic residues at specific INI-R mutations affecting native protein activity, structure or function.¹⁸ Differential effects of INI-R substitutions in IN of different HIV-1 subtypes and recombinants have been described, due to differential amino acid networks in the different IN proteins caused by polymorphic residues.¹⁸ It has been observed that the fitness cost associated with specific INI-R mutations can be different across variants. For example, subtype C IN have a low enzymatic cost associated with the G118R substitution compared with either subtype B or recombinant CRF02_AG proteins.¹⁸ This may be due to the presence of 'minor' polymorphisms and substitutions in HIV IN associated with certain subtypes and recombinants, which also can modify the genetic barrier for these variants to acquire INI-R mutations^{4,19} or to lead to alternative resistance pathways to INI-R,^{15,18} as has been reported for some HIV-1 reverse transcriptase inhibitors.²⁰ Not all HIV-1 variant-associated polymorphisms of the IN region from INI-naïve individuals may increase resistance to INI or restore the viral cost associated with certain INI-R mutations.²¹ For instance, in cell-free experiments the impacts of the H51Y and E138K substitutions on resistance and enzyme efficiency, when present with G118R, were highly dependent on viral subtype.¹⁸ The influence of HIV-1 clade on integrase mutational profiles has been recently reported in raltegravir-experienced patients.²²

Conclusions

Our study reports the natural genetic polymorphisms in the largest panel of HIV-1 variants to date from INI-naïve subjects, including groups N, O and P and group M subtypes and CRFs with available sequence in GenBank. While primary INI-R mutations are infrequent in INI-naïve individuals and rare INI-R mutations are absent, some secondary INI-R changes are frequent in an appreciable number of HIV-1 variants. The INI response in these strains might be further studied in future for a better understanding of their effect on INI susceptibility. Our data also reinforce that some newly reported INI-R changes not included in the Stanford database could be variant-associated polymorphisms more than INI-R mutations. Natural variability in INI-R positions across HIV-1 variants should be taken into consideration, as they may facilitate or delay the emergence of variants fully resistant to INIs.

Funding

This study was supported by Plan Nacional de I+D+I 2008–2011, Instituto de Salud Carlos III, FIS PI12/00240 and co-financed by the European

Variability related to integrase inhibitor resistance

Development Regional Fund 'A way to achieve Europe' (ERDF) and CIBERESP CB06/02/0053.

Transparency declarations

None to declare.

References

- Menéndez-Arias L. Molecular basis of human immunodeficiency virus type 1 drug resistance: overview and recent developments. *Antiviral Res* 2013; **98**: 93–120.
- Sharma A, Slaughter A, Jena N et al. A new class of multimerization selective inhibitors of HIV-1 integrase. *PLoS Pathog* 2014; **10**: e1004171.
- Pollicita M, Surdo M, Di Santo F et al. Comparative replication capacity of raltegravir-resistant strains and antiviral activity of the new-generation integrase inhibitor dolutegravir in human primary macrophages and lymphocytes. *J Antimicrob Chemother* 2014; **69**: 2412–9.
- Turiziani O, Montagna C, Falasca F et al. Short communication: analysis of the integrase gene from HIV type 1-positive patients living in a rural area of West Cameroon. *AIDS Res Hum Retroviruses* 2012; **28**: 1729–33.
- Lataillade M, Chiarella J, Kozal MJ. Natural polymorphism of the HIV-1 integrase gene and mutations associated with integrase inhibitor resistance. *Antivir Ther* 2007; **12**: 563–70.
- Reigadas S, Marcelin AG, Houssaini A et al. HIV-1 integrase variability and relationship with drug resistance in antiretroviral-naïve and -experienced patients with different HIV-1 subtypes. *J Antimicrob Chemother* 2013; **68**: 969–72.
- Myers RE, Pillay D. Analysis of natural sequence variation and covariation in human immunodeficiency virus type 1 integrase. *J Virol* 2008; **82**: 9228–35.
- Parczewski M, Bander D, Urbańska A et al. HIV-1 integrase resistance among antiretroviral treatment naïve and experienced patients from northwestern Poland. *BMC Infect Dis* 2012; **12**: 368.
- Bessong PO, Nwobegahay J. Genetic analysis of HIV-1 integrase sequences from treatment naïve individuals in northeastern South Africa. *Int J Mol Sci* 2013; **14**: 5013–24.
- Oliveira MF, Ramalho DB, Abreu CM et al. Genetic diversity and naturally polymorphisms in HIV type 1 integrase isolates from Maputo, Mozambique: implications for integrase inhibitors. *AIDS Res Hum Retroviruses* 2012; **28**: 1788–92.
- Vavro C, Hasan S, Madsen H et al. Prevalent polymorphisms in wild-type HIV-1 integrase are unlikely to engender drug resistance to dolutegravir (S/GSK1349572). *Antimicrob Agents Chemother* 2013; **57**: 1379–84.
- Gutiérrez C, Hernández-Novoa B, Pérez-Elías MJ et al. Prevalence of primary resistance mutations to integrase inhibitors in treatment-naïve and -experienced patients infected with B and non-B HIV-1 variants. *HIV Clin Trials* 2013; **14**: 10–6.
- Cavaco-Silva J, Abecasis A, Miranda AC et al. HIV-2 integrase polymorphisms and longitudinal genotypic analysis of HIV-2 infected patients failing a raltegravir-containing regimen. *PLoS One* 2014; **9**: e92747.
- Bellecave P, Malato L, Calmels C et al. In vitro analysis of the susceptibility of HIV-1 subtype A and CRF01_AE integrases to raltegravir. *Int J Antimicrob Agents* 2014; **44**: 168–72.
- Malet I, Fourati S, Charpentier C et al. The HIV-1 integrase G118R mutation confers raltegravir resistance to the CRF02_AG HIV-1 subtype. *J Antimicrob Chemother* 2011; **66**: 2827–30.
- Carganico A, Dupke S, Ehret R et al. New dolutegravir resistance pattern identified in a patient failing antiretroviral therapy. *J Int AIDS Soc* 2014; **17**: 19749.
- Munir S, Thierry E, Malet I et al. G118R and F121Y mutations identified in patients failing raltegravir treatment confer dolutegravir resistance. *J Antimicrob Chemotherapy* 2014; **70**: 739–49.
- Quashie PK, Oliveira M, Veres T et al. Differential effects of the G118R, H51Y and E138K resistance substitutions in different subtypes of HIV integrase. *J Virol* 2015; **89**: 3163–75.
- Maiga AI, Malet I, Soulie C et al. Genetic barriers for integrase inhibitor drug resistance in HIV type-1 B and CRF02_AG subtypes. *Antivir Ther* 2009; **14**: 123–9.
- Martinez-Cajas JL, Pai NP, Klein MB et al. Differences in resistance mutations among HIV-1 non-subtype B infections: a systematic review of evidence (1996–2008). *J Int AIDS Soc* 2009; **12**: 11.
- Wares M, Mesplède T, Quashie PK et al. The M50I polymorphic substitution in association with the R263K mutation in HIV-1 subtype B integrase increases drug resistance but does not restore viral replicative fitness. *Retrovirology* 2014; **11**: 7.
- Doyle T, Dunn DT, Ceccherini-Silberstein F et al. Integrase inhibitor (INI) genotypic resistance in treatment-naïve and raltegravir-experienced patients infected with diverse HIV clades. *J Antimicrob Chemother* 2015; doi:10.1093/jac/dkv243.

New Findings in Cleavage Sites Variability across Groups, Subtypes and Recombinants of Human Immunodeficiency Virus Type 1

Esther Torrecilla, Teresa Ll acer Delicado,  frica Holgu n*

HIV-1 Molecular Epidemiology Laboratory, Dept. of Microbiology, Hospital Ram n y Cajal- IRYCIS and CIBERESP, Madrid, Spain

Abstract

Background: Polymorphisms at cleavage sites (CS) can influence Gag and Pol proteins processing by the viral protease (PR), restore viral fitness and influence the virological outcome of specific antiretroviral drugs. However, data of HIV-1 variant-associated CS variability is scarce.

Methods: In this descriptive research, we examine the effect of HIV-1 variants on CS conservation using all 9,028 gag and 3,906 pol HIV-1 sequences deposited in GenBank, focusing on the 110 residues (10 per site) involved at 11 CS: P17/P24, P24/P2, P2/P7, P7/P1, P1/P6^{gag}, NC/TFP, TFP/P6^{pol}, P6^{pol}/PR, PR/RT^{p51}, RT^{p51}/RT^{p66} and RT^{p66}/IN. CS consensus amino acid sequences across HIV-1 groups (M, O, N, P), group M 9 subtypes and 51 circulating recombinant forms (CRF) were inferred from our alignments and compared to the HIV-1 consensus-of-consensuses sequence provided by GenBank.

Results: In all HIV-1 variants, the most conserved CS were PR/RT^{p51}, RT^{p51}/RT^{p66}, P24/P2 and RT^{p66}/IN and the least P2/P7 and P6^{pol}/PR. Conservation was significantly lower in subtypes vs. recombinants in P2/P7 and TFP/P6^{pol} and higher in P17/P24. We found a significantly higher conservation rate among Group M vs. non-M Groups HIV-1. The late processing sites at Gag (P7/P1) and GagPol precursors (PR/RT^{p51}) presented a significantly higher conservation vs. the first CS (P2/P7) in the 4 HIV-1 groups. Here we show 52 highly conserved residues across HIV-1 variants in 11 CS and the amino acid consensus sequence in each HIV-1 group and HIV-1 group M variant for each 11 CS.

Conclusions: This is the first study to describe the CS conservation level across all HIV-1 variants and 11 sites in one of the largest available sequence HIV-1 dataset. These results could help other researchers for the future design of both novel antiretroviral agents acting as maturation inhibitors as well as for vaccine targeting CS.

Citation: Torrecilla E, Ll acer Delicado T, Holgu n   (2014) New Findings in Cleavage Sites Variability across Groups, Subtypes and Recombinants of Human Immunodeficiency Virus Type 1. PLoS ONE 9(2): e88099. doi:10.1371/journal.pone.0088099

Editor: Jean-Pierre Vartanian, Institut Pasteur, France

Received: November 7, 2013; **Accepted:** January 8, 2014; **Published:** February 7, 2014

Copyright:   2014 Torrecilla et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by Ministry of Health and Social Policy through the grant of an "Independent clinical project" (EC11-131) and by research project FIS awarded by Instituto de Salud Carlos III (PI12/00240). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: africa.holgu n@salud.madrid.org

Introduction

The human immunodeficiency virus type 1 (HIV-1) Gag proteins are essential for the virus, as they have a structural and functional role in the viral cycle. They coordinate viral trafficking, membrane binding, assembly, cofactor packaging, budding, and viral modulation. Gag proteins are generated through viral maturation, essential in the viral life cycle by enabling the generation of mature infectious viral particles through the proteolytic process in specific cleavage sites (CS) of Gag precursor (Pr55^{gag}) and GagPol precursors (Pr160^{GagPol}) proteins by the viral protease (PR) [1,2]. Gag precursor is cleaved within the virion in three main structural Gag proteins: matrix (P17 or MA), capsid (P24 or CA) and nucleocapsid (P7 or NC), flanked by two spacer segments (P1 and P2) with regulatory functions [3]. Gag P6, a sixth protein of Gag precursor, plays an essential role in the release of the virus from infected cell membranes [3]. During translation of the Gag precursor an occasional ribosomal frameshift leads to the

production of a GagPol precursor protein, the abundance of which is approximately 5% that of Gag precursor [4]. GagPol precursor contains the main structural proteins matrix P17, P24, P7, a transframe protein (TFP), P6^{pol} and the three viral replication enzymes, PR, reverse transcriptase (RT) and integrase (IN) [3]. PR is activated concomitant with viral budding. As PR is only active as a dimer, it is thought that autoprocessing is initiated by dimerization of two PR domains that are embedded in the GagPol precursor [5]. Maturation triggers a second assembly event that generates a condensed conical capsid core, which organizes the viral RNA genome and viral proteins to facilitate viral replication in the next round of infection [6].

Processing of both HIV-1 Gag and GagPol polypeptides by the viral PR is highly specific, temporally regulated, and essential for the production of infectious HIV-1 particles. The differential rate of processing at each of the 11 proteolytic reactions by cleavage exists [6] and is determined by the context surrounding processing

Table 1. Gag and Pol HIV-1 proteins numbered in HXB2 genome.

		Gene		Protein	
HIV-1 proteins		Length (nucleotide)	Position (nucleotide)	Length (amino acid)	Position (amino acid)
Gag	P17	396	790–1185	132	263–394
	P24	693	1186–1878	231	395–625
	P2	42	1879–1920	14	626–639
	P7	165	1921–2085	55	640–694
	P1	48	2086–2133	16	695–710
	P6	156	2134–2289	52	711–762
	Total	1500	790–2289	500	263–762
Pol	PR	297	2253–2549	99	751–849
	RT ^{p51}	1320	2550–3869	440	850–1289
	RT ^{p66}	360	3870–4229	120	1290–1409
	IN	864	4230–5093	288	1410–1697
	Total	2841	2253–5093	947	751–1697

Nucleotides and amino acids numbered according to HXB2 subtype B reference strain (GenBank accession number K03455). P17, matrix; P24, capsid; P2, spacer peptide 1; P7, nucleocapsid; P1, spacer peptide 2; PR, protease; RT, reverse transcriptase; IN, integrase; TFP, transframe protein. Retrieved from <http://www.hiv.lanl.gov/>. doi:10.1371/journal.pone.0088099.t001

sites of the CS [7]. However, the precise mechanisms governing the rates of the cleavage events are still not fully understood [7].

The physical consequence of Gag cleavage is a morphological rearrangement of the non-infectious immature particle to a mature infectious particle. For this reason, amino acid substitutions on Gag proteins, included in CS, could influence processing [2,8], morphogenesis, budding [9], the virus replicative capacity or *viral fitness* [3,10] and the virological outcome of specific regimens, particularly to protease inhibitors (PI) [5,11–20]. In fact, several Gag substrate mutations, included in CS, can confer PI resistance in the absence and/or presence of PR mutations [17–20]. The fundamental role of proteolytic maturation in the generation of infectious particles makes inhibition of this process an attractive target for therapeutic intervention. Thus, a new class of potential antiretroviral drugs targeting individual Gag CS has entered development [21].

Whether or not the processing regulation is different across HIV-1 variants remains unclear. It is well known that HIV-1 shows a high genetic diversity due to its high replication rate, the error-prone RT and the recombination events between HIV-1 variants occurring during the viral replication after co-infection and/or superinfection events [22–24]. A large number of HIV-1 variants have been described based on viral sequences homology and HIV-1 has been divided into four groups: M (main), O (outlier), N (non-M, non-O) and P [23]. HIV-1 Group M is subdivided into 9 subtypes (A, D, F, H, J, K), at least 58 circulating recombinant forms (CRF) (<http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>)-designated by a number and the genetic subtypes present in their genome- and multiple unique recombinant forms (URF), widely spread throughout the world and with different recombination breakpoints from those found in CRFs. At least 20% of the 34 million infected humans have an inter-subtype URF or CRF [25] and new inter-subtype recombinants have increasing prevalence and complexity in the pandemic, including in some European countries [26]. Genetic variability in PR and CS provide the potential to modulate PR activity and susceptibility to PI [20]. For instance, CS polymorphisms in certain HIV-1 group M variants can influence the virological outcome of a first-line LPV/r single drug regimen [19].

Despite the high biological relevance of CS during HIV-1 maturation and the importance of the knowledge of CS conservation for the design of both novel antiretroviral agents acting as maturation inhibitors as well as for vaccine targeting CS in future, scarce data of HIV-1 variant-associated CS variability is available. Previous reports only analyzed a limited number of HIV-1 variants and site sequences [3,27,28]. Thus, the goal of our descriptive analysis was to analyze, for the first time, the conservation rate at amino acid level of each individual protease CS located within Gag or Pol for all HIV-1 groups, Group M subtypes and recombinants circulating in the HIV/AIDS pandemic. For this purpose we used a large dataset of HIV-1 sequences routinely deposited at Los Alamos National Center for Biotechnology Information or GenBank. We also defined the consensus sequences at each CS in all HIV-1 variant, identifying the highly conserved amino acids residues in each CS.

Methods

Sequence Data

All the available HIV-1 *gag/pol* sequences were retrieved from GenBank, (<http://www.ncbi.nlm.nih.gov/>). The 12,934 *gag/pol* sequences comprised 2,844 nucleotides, located from 790 to 2,292 in *gag* and from 2,253 to 5,096 in *pol* encoding the proteins shown in Table 1. These sequences belonged to 4 groups (M, O, N, P), 9 Group M subtypes (A: sub-subtypes A1 and A2, B, C, D, F: sub-subtypes F1 and F2, G, H, J and K), 51 of the 58 CRF currently described, and with available sequences at GenBank and URF (Figure 1). For the subsequent analysis, we grouped in 12 recombinant families the closely related CRF sharing the same parental subtypes and very similar recombination patterns (Figure 2), as previously recommended [23]. All *gag/pol* nucleotides sequences were retrieved in FASTA format, including the subtype B HXB2 reference sequence. The MEGA (Molecular Evolutionary Genetics Analysis. Arizona States University, Tempe) program version 5.05 (<http://www.megasoftware.net/>) [29] was used to perform the nucleotides alignments and to translate them into amino acids.

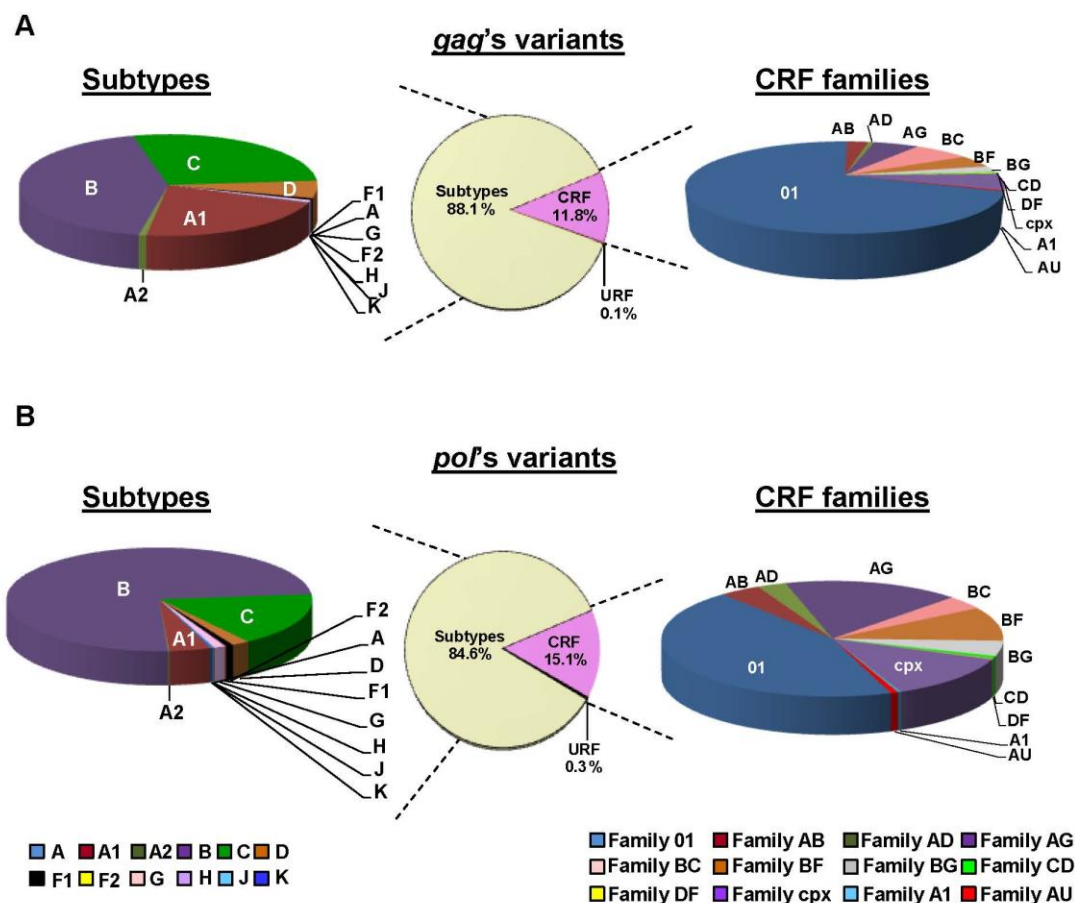


Figure 1. Distribution of HIV-1 Group M subtypes and CRF families. A total of 12,848 HIV-1 Group M sequences were retrieved from GenBank: 8,985 *gag* (A) and 3,863 *pol* (B) sequences.
doi:10.1371/journal.pone.0088099.g001

Identification of *gag* and *pol* Coding Regions and CS Sequences Defined at GenBank

After performing the alignments, we determined the residues and their location in Gag and Pol proteins (**Table 1**), identifying their nucleotides and amino acids and numbering them according to HXB2 subtype B reference strain (GenBank accession number K03455). We then identified the residues and the location of 11 cleavage sites (CS) within Gag and GagPol precursors: P17/P24, P24/P2, P2/P7, P7/P1, P1/P6^{gag}, P7/TFP, TFP/P6^{pol}, P6^{pol}/PR, PR/RT^{p51}, RT^{p51}/RT^{p66} and RT^{p66}/IN according to HXB2 sequence.

Inferred Consensus Sequences

The consensus sequence is considered the sequence carrying the most frequent residues, either nucleotides or amino acids, at each position in a multiple sequence alignment. We collected all Gag and Pol consensus sequences available in GenBank (<http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>). The HIV-1 Group M variants with inferred consensus sequences in GenBank are indicated in **Figure 2**, and were calculated as explained in <http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html#consensus>. Using our amino acid alignment, composed of 12,934 sequences, we determined new

consensus sequences for each HIV-1 group and each Group M subtype, CRF and URF in the 11 CS (**Figures 3 and 4**). Then, we manually compared our inferred variant-associated consensus sequences at each CS with the ones provided by GenBank when available, showing the discrepancies.

We also retrieved the consensus-of-consensuses sequence provided by GenBank in order to generate an alignment of *gag* and *pol* individual consensus sequences that were used to analyze the conservation rate across sites and HIV-1 variants (**Figure 5**).

Amino Acid Conservation Rate at CS Across HIV-1 Variants

All *gag* and *pol* sequences from GenBank were grouped according to the HIV-1 variant. We manually compared the degree of amino acid conservation in each CS, determined by the number of coincident amino acids among the 10 residues of each CS, in all downloaded sequences from each given variant with respect to the consensus-of-consensuses sequence provided by GenBank. The exact percentage of conserved amino acid residues for each HIV-1 variant and site with respect to the GenBank consensus-of-consensuses amino acid sequence was calculated counting the number of coincident residues in each of the 10 positions in the site in all sequences ascribed to a given variant

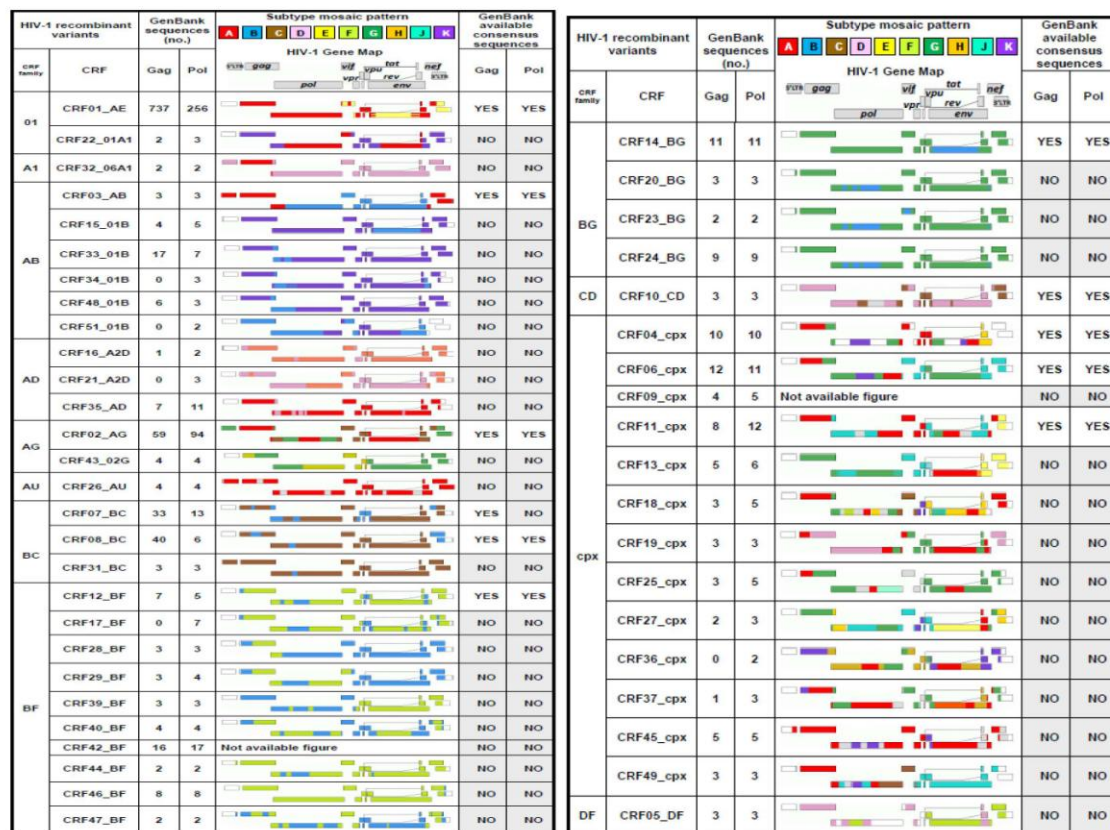


Figure 2. Gag and Pol HIV-1 recombinants sequences grouped by families. Availability of consensus sequences at GenBank. CRF sequences were grouped in 12 recombinant families; no, number; CRF, circulating recombinant forms <http://www.hiv.lanl.gov/content/sequence/HIV/CRFs/CRFs.html>; URF, unique recombinant forms. Other variants with consensus sequences from GenBank were: A1, A2, B, C, D, F1, G, H and K subtypes for *gag* and: A1, A2, B, C, D, F1, F2, G and H subtypes for *pol*. <http://www.hiv.lanl.gov/content/sequence/NEWALIGN/align.html>. doi:10.1371/journal.pone.0088099.g002

divided by the total number of retrieved sequences for each variant and multiplied by the 10 residues of the site, expressing the result in percentages. To clarify results, we established a color code to show the conservation level in each site and HIV-1 variant (Figure 5).

Data Analysis

Changes in rates were assessed using the chi-square analysis. Statistical analyses were performed using Epi Info v6.0 (Centers for Disease Control and Prevention, Atlanta, GA, USA). Significance was set at $p < 0.05$.

Results

Gag/Pol HIV-1 Sequences Used for the Analysis and Variants Distribution

A total of 9,028 Gag and 3,906 Pol HIV-1 sequences were downloaded from GenBank database. They included 43/43 Gag/Pol sequences from 3 HIV-1 Groups (O, N, P) and 8,985/3,863 Gag/Pol sequences from Group M. Figure 1 shows the Group M variants distribution of our retrieved sequence dataset, including a total of 7,913/3,269 Gag/Pol sequences from 9 HIV-1 group M subtypes (A: subsubtypes A1 and A2, B, C, D, F: subsubtypes F1

and F2, G, H, J and K), 1,060/583 Gag/Pol sequences ascribed to 51 CRF and 12/11 Gag/Pol URF sequences.

In order to simplify the analysis, we grouped all the sequences from the 51 CRFs in 12 different CRF families according to a similar recombination pattern (Figure 2). The downloaded sequences for each subtype and CRF family are detailed in Figure 5. Despite the large difference in the number of 8,985 Gag/3,863 Pol retrieved sequences, the specific distribution of HIV-1 Group M subtypes and CRF families was similar for both genes (Figure 1). Recombinants displayed 11.9% *gag* and 15.4% *pol* sequences. Among subtypes, sequences from subtype B were the most represented in both *gag/pol* (43%/74.3%) coding regions, followed by sequences ascribed to subtype C (27.4%/16.7%), sub-subtype A1 (21.6%/4.3%) and subtype D (5.6%/1.9%). There were no *gag* sequences from sub-subtype F2 and subtypes J and K available in our dataset. Within the recombinants, family 01 (69.7%/44.4%) was the most represented, followed by families BC (7.2%), AG (5.9%), cpx (5.6%) and BF (4.5%) in *gag* and by families AG (16.8%), cpx (12.5%), BF (9.4%), and BG (4.3%) in *pol*, among others. URF sequences represented less than 0.3% of downloaded sequences (12 *gag* and 11 *pol* sequences).

Cleavage Sites Variability across HIV-1 Variants

HIV-1 variants in CS1		P17					P24				
		V	S	Q	N	Y	P	I	V	Q	N
Group M	02_AG, 32_06A1, 43_02G	S
	06_cpx	L
	14_BG	A
	27_cpx	VA	QH	.
	45_cpx, 49_cpx	I
Group N*		I	.	R	.	.	.	L	.	.	.
Group O*		T	gs	V	.	#1	.
#1, apsq.											

HIV-1 variants in CS2		P24					P2				
		K	A	R	V	L	A	E	A	M	S
Group M	F2*, URF*	.	.	.	I
	35_AD*	G
	Group P* & O*	A

HIV-1 variants in CS3		P2					P7				
		T	T	—	I	M	M	Q	R	G	N
Group M	A*, A1, J, 35_AD*	.	N	—	—
	A2	.	N	—	#1
	B	S	A	—	T
	C	#2	N	—	—	S	.
	D	#3	A	—	A
	F1	.	#4	—	—	V	.	.	.	K	S
	F2*	TA	A	—	—	K	S
	G, 04_cpx, 13_cpx*	A	A	A	A	K	S
	H	A	N	AT	A	K	.
	K	PS	AV	—	—	V	.	.	.	RK	.
	01_AE	A	N	—	—
	02_AG	S	N	—	—
	03_AB, 32_06A1*	A	N	—	—	K	S
	05_DF*	S	#5	—	A	V
	06_cpx	#6	A	A	A	K	S
	07_BC	A	—	—	—	L	.	.	.	S	.
	08_BC	.	—	—	—	L	.	.	.	S	.
	09_cpx*	.	N	—	—	L
	10_CD	G	N	—	A
	11_cpx, 31_BC*	.	N	—	—	S	.
	12_BF, 46_BF*	.	.	—	—	V	.	.	.	K	S
	14_BG	A	.	—	—	K	S
	15_01B*	A	N	—	—	IV	GD
	16_A2D*	.	N	S	N	K	.
	18_cpx*	#7	N	—	—
	19_cpx*	S	#8	A	A	K	S
	20_BG*, 24_BG*	A	G	A	S	K	S
	22_01A1*, 33_01B*, 48_01B*	A	G	—	—
	23_BG*	A	G	A	S	ML	.	.	.	RK	S
	26_AU*	.	N	—	—
	27_cpx*	A	AV	AT	A	K	SG
	28_BF*	A	A	—	—	V	.	.	.	K	S
	29_BF*	.	#9	—	—	V	.	.	.	K	S
	37_cpx*	P	.	—	—	K	Q
	39_BF*	Q	.	—	N
	40_BF*	.	A	—	—	V	.	.	.	K	S
	42_BF*	.	A	—	—	V	L	.	.	.	S
	43_02G*, 25_cpx*	A	A	A	A	K	.
	44_BF*	.	na	—	—	V	.	.	.	RK	S
	45_cpx*	T	N	—	—
	47_BF*	S	N	—	—	V	.	.	.	K	.
	49_cpx*	#10	#11	—	—
	URF*	.	#12	—	—
Group P*		S	K	.	.	R	R	V	YK	S	Q
Group N*		.	.	—	—	S	V	F	A	.	.
Group O*		#13	.	—	—	A	V	F	.	.	Q

#1, tnla; #2, tainvgkmpqier; #3, aglvsetpm; #4, tsani; #5, pat; #6, tang; #7, tra; #8, nhs; #9, nta; #10, gps; #11, shn; #12, antv; #13, elk.

HIV-1 variants in CS4		P7					P1				
		E	R	Q	A	N	F	L	G	K	I
Group M	03_AB	R	.
	05_DF*	V
	14_BG	.	S	K
	20_BG*, 23_BG*, 48_01B*	L
	37_cpx*	R	V
Group P*		G
Group N*		G	G
Group O*		G	K	Y

HIV-1 variants in CS5		P1					P6 ^{99g}				
		R	P	G	N	F	L	Q	S	R	P
Group M	A*, A1, A2, F1, F2*, 06_cpx, 25_cpx*, 49_cpx*, 12_BF, 28_BF*, 29_BF*, 42_BF*, 46_BF*, 47_BF*, 14_BG, 32_06A1*	N	.	.
	G	N	.	#1
	01_AE, 02_AG, 22_01A1*, 09_cpx*, 45_cpx*, 37_cpx*, 03_AB, 35_AD*
	15_01B*	PL
	16_A2D*	T
	18_cpx*	K	.
	23_BG*, 26_AU*	N	.	LI
	24_BG*	N	.	L
	40_BF*	N	.	PL
	44_BF*	N	.	PL
	48_01B*	I	.	.
Group P*		Y	.	.	K	Q	V
Group N*		T	T	K
Group O*		Y	#2	.	K	Q	V

#1, pti; #2, vami.

Figure 3. HIV-1 variants showing differences in CS1–CS5 amino acid vs. consensus-of-consensuses sequence from GenBank. Changes are only indicated when they appeared in a specific position in at least 50% of the GenBank downloaded sequences in order to compare them with the GenBank consensus-of-consensuses sequence. Asterisks indicate the HIV-1 variants shown in **Figure 2** with non available consensus sequence in GenBank. Black represents highly conserved amino acid residues and present in more than 99% of the 9,028 Gag and 3,906 GagPol HIV-1 sequences with respect to the consensus-of-consensuses sequence. When two residues within the analyzed sequences from each HIV-1 variant

showed a conservation of 50% the two code letters were written in the upper case. When 3 or more residues appear in the same position and none presented a conservation of more than 50%, they were shown in lower case letters, which represented higher to lower conservation.
doi:10.1371/journal.pone.0088099.g003

HIV-1 Variant-specific *gag/pol* Consensus Sequences Available at GenBank

Figure 2 shows the specific subtypes and recombinants with consensus sequences in *gag* and *pol* available in GenBank, which carries the most frequent residue, either nucleotide or amino acid, at each position in a multiple sequence alignment. **Table 2** summarizes the amino acids involved in each of the 11 CS (10 amino acids per site) in the HXB2 isolate as well as the consensus-of-consensuses sequence for each CS, defined by GenBank after the alignment of 28 *gag/24 pol* individual consensus sequences, corresponding to 8/7 subtypes among 9 in Group M and to 11/10 CRF within the 58 described (**Figure 2**). The consensus-of-consensuses sequence was taken as reference for the analysis of the conservation at amino acid level across variants in the 110 residues (10 amino acids in each of the 11 CS), as described in Methods.

New Inferred Consensus Sequence in HIV-1 Groups, Subtypes and Recombinant vs. that Provided by GenBank

Since *gag* and *pol* consensus sequences were not defined by GenBank in all HIV-1 subtypes and CRF, we deduced our personal consensus sequence for all HIV-1 variants using our generated alignment of 9,028 Gag and 3,906 Pol HIV-1 sequences. We determined that the rate of amino acid residues among the retrieved sequences coincided with the consensus-of-consensuses in the corresponding site. For the first time, we inferred the consensus sequence in each site for the different HIV-1 groups and for all subtypes, sub-subtypes and recombinants within Group M. **Figures 3 and 4** show the HIV-1 variants that carry amino acid differences with the corresponding consensus-of-consensuses sequence from GenBank in CS. We identified when our inferred consensus sequence presented the same amino acid residue as consensus-of-consensuses provided by GenBank. All discrepancies found between our inferred variant-specific CS consensus sequences with the consensus-of-consensuses provided by GenBank were also identified (see **Table S1**).

Identification of Highly Conserved Residues at CS

Interestingly, we identified 52/110 (47.3%) amino acids conserved in more than 99% of the 9,028 Gag and 3,906 Pol HIV-1 sequences with respect to the consensus-of-consensuses sequence and these are marked in black in **Figures 3 and 4**. Thus, nearly half of the residues involved in the 11 CS could accept a different degree of variant-dependent variability. Among sites, PR/RT^{P51} presented the highest number of highly conserved residues (9/10), followed by RT^{P51}/RT^{P66} and P24/P2 (7/10), RT^{P66}/IN (6/10), P7/P1 and P7/TFP (5/10), P1/P6^{gag} and P6^{pol}/PR (4/10), P17/P24 (3/10) and TFP/P6^{pol} and P2/P7 (1/10).

Observed Differences in CS Conservation Rates Across HIV-1 Variants and Sites

We evaluated the percentage of conserved residues in the retrieved sequences for each HIV-1 variant and site with respect to the GenBank consensus-of-consensuses amino acid sequence, as explained in Methods. We established a color code to indicate the different levels of conservation, and the exact amino acid conservation rate in each CS and variant (**Figure 5**). Interestingly, despite the structural and functional roles of proteins in the viral

cycle, we observed different conservation rates across the sites and HIV-1 variants.

Conservation Among CS

In all HIV-1 variants, including all sequences ascribed to the 4 groups, we defined the conservation rate in each site (**Figure 5**). The CS with the highest number of conserved residues were CS9 (PR/RT^{P51}, 99%), CS10 (RT^{P51}/RT^{P66}, 98%), CS2 (P24/P2, 98%), CS11 (RT^{P66}/IN, 97%), CS1 (P17/P24, 96%), CS4 (P7/P1, 96%) and CS6 (P7/TFP, 96%). The least conserved CS across HIV-1 groups, Group M subtypes and recombinants were CS3 (P2/P7, 71%), CS8 (P6^{pol}/PR, 80%), CS7 (TFP/P6^{pol}, 81%) and CS5 (P1/P6^{gag}, 89%). CS8 and CS3 showed different lengths across variants (data not shown). We observed a significantly higher conservation at the last processing sites in Gag (CS4, P7/P1) and GagPol (CS9, PR/RT^{P51}) precursors compared to the first processing site (CS3, P2/P7) in the 4 HIV-1 groups (**Figure 6**), according to the processing order previously defined [3,5,30].

Conservation among HIV-1 Groups

We observed differences in the CS conservation rate across HIV-1 groups and sites (**Figure 5**). Interestingly, CS10 (RT^{P51}/RT^{P66}) showed more than 90% of conserved residues regarding consensus-of-consensuses amino acid sequence in the 4 HIV-1 groups. Comparing M and non-M Groups, we observed higher conservation in CS9 (99% and 98%, respectively) and in CS10 (98% and 91%, respectively). However, CS7 (TFP/P6^{pol}) presented the poorest conservation rate across non-M Groups (41%), followed by CS5 (P1/P6^{gag}, 54%), CS3 (P2/P7, 56%), CS8 (P6^{pol}/PR, 60%) and CS1 (P17/P24, 68%). Group O showed the lowest conservation in 6 of the 11 CS (CS1, CS2, CS4, CS6, CS8 and CS11), Group N in CS1 and CS7 and Group P in CS3, CS5, CS9 and CS10 (**Figures 3 and 4**). Considering the 11 CS, we found a significantly higher conservation rate among Group M vs. non-M Groups HIV-1 variants (91% vs. 71%, $p < 0.001$), probably due to the use of group M consensus for comparison.

Conservation among Group M Subtypes and Recombinants

Seven sites (CS1, CS2, CS4, CS6, CS9, CS10 and CS11) were well conserved within the total HIV-1 Group M subtypes and recombinants, showing more than 90% conservation (**Figure 5**). Four sites (CS3, CS5, CS7 and CS8) were more variable in a large number of HIV-1 variants. The lowest conservation rate in the 11 CS was found in the following HIV-1 Group M subtypes and recombinants: CS1 (P17/P24) in sub-subtype A2 (89%) and in AG recombinant family (77%); CS2 (P24/P2) in sub-subtype F2 (90%) and recombinant families AD and URF (both 94%); CS3 (P2/P7) in subtypes G (52%) and recombinant family DF (53%); CS4 (P7/P1) in subtype G (84%) and recombinant family DF (87%); CS5 (P1/P6^{gag}) in sub-subtype A1 (76%) and recombinant family AU (73%); CS6 (P7/TFP) in subtype G (84%) and recombinant family BG (88%); CS7 (TFP/P6^{pol}) in subtype B (77%) and recombinant families AB and BC (both 75%); CS8 (P6^{pol}/PR) in sub-subtype A2 and subtype C (both 68%) and recombinant family BG (67%); CS9 (PR/RT^{P51}) sub-subtype A2 (90%) and recombinant family A1 (95%); CS10 (RT^{P51}/RT^{P66}) in clades H (88%) and G (89%) and recombinant family A1 (90%); and CS11 (RT^{P66}/IN) in sub-subtype F1 (90%) and recombinant family BF (91%). Thus,

Cleavage Sites Variability across HIV-1 Variants

HIV-1 variants in CS6		P7					TFP				
		E	R	Q	A	N	F	F	R	E	N
Group M	J*	D
	05_DF*	S
	08_BC	I
	14_BG	.	S	K	D
	15_01B*	#1
	20_BG*	T
	23_BG*	T
	24_BG*	T
	33_01B*	A
	40_BF*	ND
	48_01B*	K	#2
Group P*		G	EK	D
Group N*		G	#3	G
Group O*		G	K	EQ	I

#1, ntd; #2, adt; #3, ekq.

HIV-1 variants in CS8		pg ^{pol}					PR				
		V	—	S	L	S	F	P	Q	I	T
Group M	A*	A	#1	.	F	.	FL
	A1	#2	#3	T	F
	A2	.	H	.	C	N
	B, D, 49_cpx*, 35_AD*	.	.	.	F
	C	G	—	T	.	N
	F1	.	P
	F2	VG	S	.	D
	G	#4	—	.	.	L
	H	—	—	.	N
	J*	.	.	.	S	N
Group M	K*	EG	SP	.	F	N
	01_AE	S	S	.	F
	02_AG	I	S	.	F	N
	03_AB	A	S	.	F	N
	04_cpx, 11_cpx, 07_BC*	I	—	.	F	N
	05_DF*, 12_BF*, 17_BF*, 29_BF*	.	P
	06_cpx, 32_06A1*	I	—
	08_BC, 31_BC*	—	—	T	.	N
	09_cpx*	#5	P	.	F
	13_cpx*	I	—	.	F	N	C
Group M	14_BG	I	—	P	.	L
	15_01B*	#6	S	.	F
	16_A2D*	—	—	.	C	N
	18_cpx*	I	—	.	F	L
	19_cpx*	—	—	.	N	L
	20_BG*, 23_BG*, 24_BG*	—	—	G	.	N	L
	22_01A1*	A	S	.	F	.	L
	25_cpx*	I	—	.	#7	#8
	26_AU*	VA	—	.	FS
	27_cpx*	—	—	.	F
Group M	28_BF*	.	sp	.	.	#9
	33_01B*, 34_01B*, 48_01B*	I	—	.	F
	36_cpx*	I	S	.	.	N
	37_cpx*	I	P	.	F
	39_BF*, 21_A2D*	.	—	.	.	N
	40_BF*	.	P	.	.	SN
	42_BF*	.	P	.	.	N
	43_02G*	T	—	.	I	.	L
	44_BF*	.	P	.	.	SC
	45_cpx*	ag	P
Group M	46_BF*	#10	P
	47_BF*	VG	TS
	URF*	G	S	.	F	N
	Group P*	GV	IV	P	F	SN	L
	Group N*	.	P	T	.	N
	Group O*	L	—	.	V	C	L	.	.	.	P

#1, lgi; #2, gaetdvslm; #3, pslyfghirt; #4, ltn; #5, atv; #6, yes; #7, ycfi; #8, gdc; 9, sng; #10, lqiv

HIV-1 variants in CS7		TFP					P6 ^{pol}				
		E	N	L	A	F	Q	Q	G	E	A
Group M	A*, H, 09_cpx*, 13_cpx*, 25_cpx*, 45_cpx*, 03_AB	R	.	.
	B, D	P	.	.	K	.
	C, 29_BF*, 43_02G*, 31_BC*, 28_BF*	P
	J*	.	D	R	.	.
	01_AE	K	.
	04_cpx	.	.	V	R	.	.
	05_DF*	.	S	.	.	.	P
	07_BC*	L	P	.	.	K	.
	08_BC	.	I	.	.	.	P
	10_CD	R	K	.
Group M	15_01B*	.	#1	.	.	.	QP	.	.	K	.
	20_BG*, 24_BG*	.	T	.	.	.	P	.	.	K	.
	23_BG*	.	T	.	.	.	P	.	.	EK	.
	33_01B*	.	A	.	.	.	P
	34_01B*	.	I	K	.
	40_BF*	.	ND	.	.	FS	PL
	42_BF*, 21_A2D*	#2
	48_01B*	K	#3	.	.	.	P
	51_01B*	.	D	.	.	.	P	.	.	K	.
	Group P*	EK	D	.	.	S	WG	G	Q	.	.
Group N*		#4	G	.	V	S	L	.	R	.	T
Group O*		EQ	I	.	.	S	G	G	H	.	.

#1, ntd; #2, lq; #3, adt; #4, ekq.

HIV-1 variants in CS9		PR					RT ^{P51}				
		C	T	L	N	F	P	I	S	P	I
Group M	11_cpx	IV
	16_A2D	IV	.	PT	.
	32_06A1	FL
Group P*		.	.	.	S	PS	.

HIV-1 variants in CS10		RT ^{P51}					RT ^{P66}				
		G	A	E	T	F	Y	V	D	G	A
Group M	G, H, 14_BG, 13_cpx	Y
	07_BC*, 08_BC, 32_06A1*	.	V
	17_BF*	E
	26_AU*	FY
	36_cpx*	.	.	.	EG	TR
	43_02G*	.	AV
	Group P*	.	IT
	Group N* & O*	Y

HIV-1 variants in CS11		RT ^{P66}					IN				
		I	R	K	V	L	F	L	D	G	I
Group M	F1, 36_cpx*, 05_DF*, 17_BF*, 28_BF*, 29_BF*, 42_BF*, 46_BF*, 12_BF
	03_AB	.	.	#1
	23_BG*	.	.	KR
	51_01B*	.	.	R
	Group P* & O*	.	.	R	E	.	.

#1, kre.

Figure 4. HIV-1 variants showing differences in CS6-CS11 amino acid vs. consensus-of-consensuses sequence from GenBank. See legend of **Figure 3**.
doi:10.1371/journal.pone.0088099.g004

subtype G showed the highest variability in CS3, CS4 and CS6 and subtype B in CS7 compared to other Group M subtypes. The recombinant families DF, BG, A1 and BF showed the highest variability in CS3 (**Figure 5**). The CS2, CS6, CS9, CS10 and CS11 were highly conserved (94%–99%) across all CRF families and URF. Overall, recombinants showed the same conservation as subtypes (91%) in the 11 analyzed CS, the BG recombinant family with 87%, subtype G with 86%, sub-subtype A2 with 88% and subtype H with 89%. The HIV-1 variants that presented the lowest CS conservation were significantly lower in subtypes *vs.* recombinants in CS3 (70% *vs.* 76%, $p<0.001$) and in

CS7 (80% *vs.* 87%, $p<0.001$) and higher in CS1 (97% *vs.* 90%, $p<0.001$).

Conservation among Sub-subtypes in Specific Sites

Sub-subtype A2 presented significantly lower conservation than sub-subtype A1 in CS1 (89% *vs.* 98%, $p<0.001$), in CS3 (70% *vs.* 85%, $p<0.001$) and in CS9 (90% *vs.* 99%, $p<0.001$), significantly higher conservation in CS5 (80% *vs.* 76%, $p=0.04$) and the conservation was of great significance in CS11 (100% *vs.* 92%, $p=0.06$). Sub-subtype F2 showed superior conservation compared

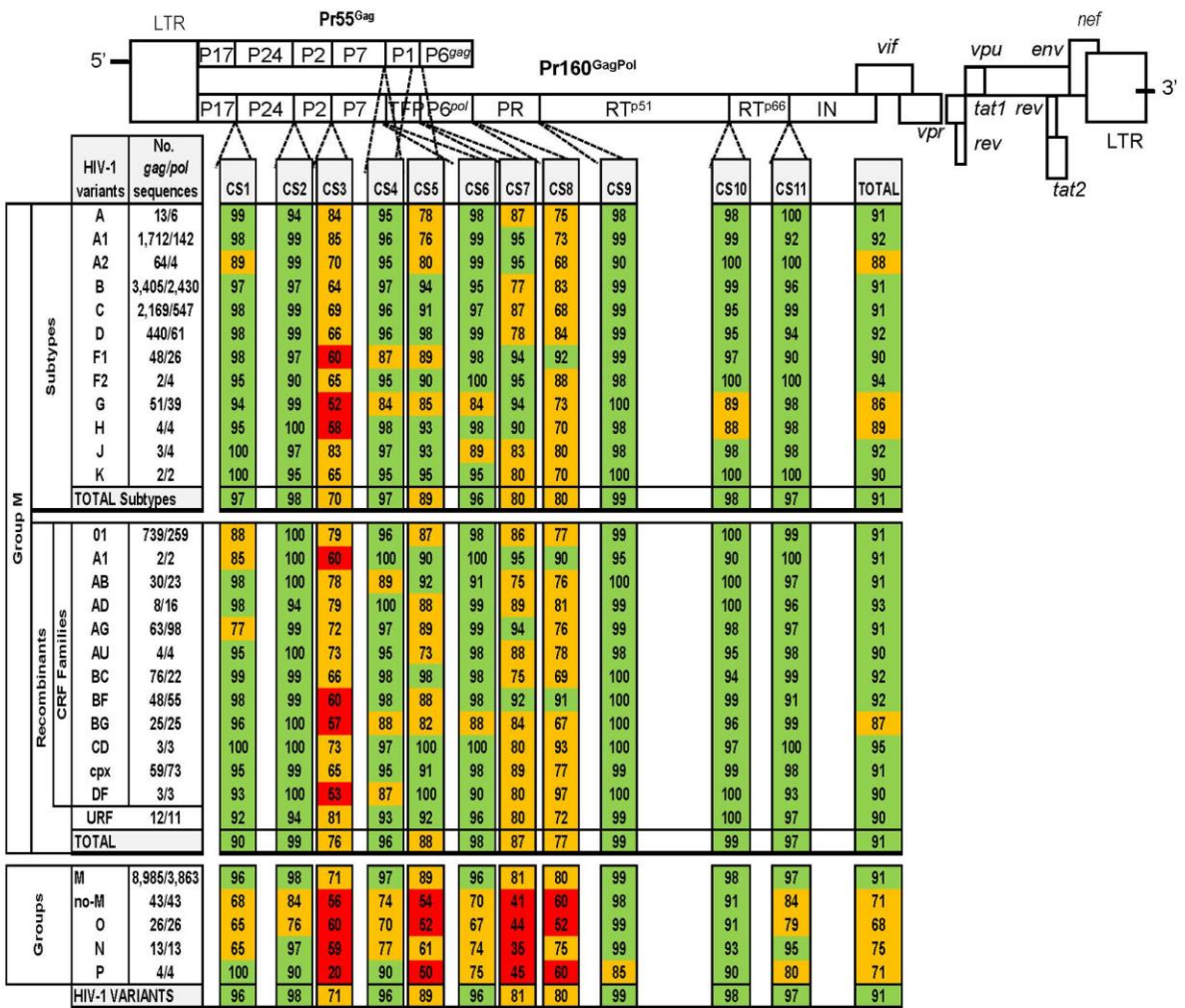


Figure 5. Amino acids CS conservation located in Gag and GagPol precursors in all HIV-1 variants. The conservation was determined by comparing our inferred consensus sequences with sequences from each HIV-1 variant *vs.* consensus-of-consensuses sequence retrieved from GenBank. Coloured boxes indicate the CS conservation rate at amino acid level: green ($\geq 90\%$), orange (51–89%) and red ($\leq 50\%$). The number in each coloured box shows the rate of conserved amino acid in each CS in all sequences of the corresponding HIV-1 variant. CS, cleavage site; P17, matrix; P24, capsid; P2, spacer peptide 1; P7, nucleocapsid; P1, space peptide 2; TFP, transframe protein; PR, protease; RT, retrotranscriptase; IN, integrase; CRF, circulant recombinant form; URF, unique recombinant form.
doi:10.1371/journal.pone.0088099.g005

Table 2. Gag and Pol HIV-1 CS numbered in HXB2 genome.

HIV-1		Gene				Protein	
No.	Name	Consensus-of-consensus sequence from GenBank	HXB2 sequence	Length (nucleotide)	Position (nucleotide)	Length (amino acid)	Position (amino acid)
#1	P17/P24	VSQNY/PIVQN	VSQNY/PIVQN	30	1171–1200	10	390–399
#2	P24/P2	KARVL/AEAMS	KARVL/AEAMS	30	1864–1893	10	621–630
#3	P2/P7	TT-IM/MQRGN	SATIM/MQRGN	30	1906–1935	10	635–644
#4	P7/P1	ERQAN/FLGKI	ERQAN/FLGKI	30	2071–2100	10	690–699
#5	P1/P6 ^{gag}	RPGNF/LQSRP	RPGNF/LQSRP	30	2119–2148	10	706–715
#6	P7/TFP	ERQAN/FFREN	ERQAN/FFRED	30	2071–2100	10	690–699
#7	TFP/P6 ^{pol}	ENLAF/QQGEA	EDLAF/LQGKA	30	2094–2123	10	698–707
#8	P6 ^{pol} /PR	VLSLF/PQITL	VSFNF/PQVTL	30	2238–2267	10	746–755
#9	PR/RT ^{P51}	CTLNF/PISPI	CTLNF/PISPI	30	2535–2564	10	845–854
#10	RT ^{P51} /RT ^{P66}	GAETF/YVDGA	GAETF/YVDGA	30	3855–3884	10	1285–1294
#11	RT ^{P66} /IN	IRKVL/FLDGI	IRKVL/FLDGI	30	4215–4244	10	1405–1414

Nucleotides and amino acids numbered according to HXB2 subtype B reference strain (GenBank accession number K03455). Pr160^{GagPol} includes CS #1 to 11 and Pr55^{Gag} includes CS #1 to 5. Underlined amino acids show the changes in the HXB2 sequence vs. the consensus-of-consensus sequence from GenBank. P17, matrix; P24, capsid; P2, spacer peptide 1; P7, nucleocapsid; P1, spacer peptide 2; PR, protease; RT, reverse transcriptase; IN, integrase; TFP, transframe protein; No., CS position in Gag and GagPol precursors; CS, cleavage site. Retrieved from <http://www.hiv.lanl.gov/>. doi:10.1371/journal.pone.0088099.t002

to sub-subtype F1 in CS11 (100% vs. 90%, $p = 0.04$), although the number of available F2 sequences was very low (see **Figure 5**).

Discussion

HIV-1 genomes analysis provides useful biological information in terms of the structure and function of viral proteins [31]. The correct core formation is essential for the production of infectious HIV particles and this is known to be dependent on accurate proteolytic processing of Gag. Thus, mutations that disrupt the cleavage of individual sites or alter the order in which sites are cleaved result in aberrant particles that have significantly reduced infectivity [6]. Although other publications previously reported that certain CS were more conserved than others, they only analyzed a very limited number of HIV-1 variants and site sequences [3,27,28]. Thus, to our knowledge, our study is the first

to evaluate the conservation rate in 11 CS within Gag and GagPol precursors and to define the consensus sequences in each site using a large sequence dataset including all Group M subtypes and most CRF. Furthermore, it is the first study that includes sequences from Groups N, O and P, identifying completely conserved residues at CS present in all 4 groups. We showed the conservation rate in each HIV-1 variant and CS, finding different conservation rates across sites in the 4 HIV-1 groups and in Group M variants, including a complete panel of recombinants, whose prevalence and complexity is increasing in the pandemic [23]. In fact, the different clade distribution for *gag* and *pol* sequences retrieved for GenBank used in the study could be explained by the large number of recombinants circulating in pandemic, with different clades in different viral genome genes.

New Findings on CS Variability Across HIV-1 Variants

Only a limited number of studies have previously evaluated the natural variation within *gag* and CS [3,5,12,28,32,33]. However, these have mainly focused on subtypes B and/or C and they have analyzed a smaller dataset or a limited number of CS in most cases. Furthermore, the majority of the studies used HXB2 subtype B as reference sequence for conservation analysis [5,32,33], pNL-4-3 subtype B [12] or the Group M most recent common ancestor sequence [3]. Only one used the consensus-of-consensus sequence provided by GenBank as a reference for comparisons [28]. None inferred a consensus sequences for each analyzed HIV-1 variant and site. Other studies included either recombinants or non-M Group sequences. Despite the wide variety in the number of sequences that we downloaded from GenBank for Group M (8,985/3,863 *gag/pol* sequences) with respect to the rest of groups (43/43 *gag/pol* sequences) or certain subtypes (H, J, K), sub-subtypes (F2) or CRF, available data permitted the establishment of a comparison among conservation rates at CS and we were able to define specific-variant differences at each CS consensus sequence for each HIV-1 group, subtype, CRF and URF (see **Figure 5**). Our data reflects that the degree of conservation differs between individual amino acid positions at CS

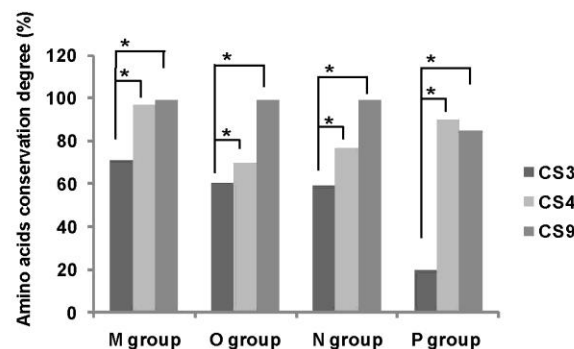


Figure 6. Conservation of the first and late processing sites at Gag and GagPol precursors. Late processing sites at Gag (CS4, P7/P1) and GagPol (CS9, PR/RT^{P51}) precursors and first CS site (CS3, P2/P7) according to the CS order previously described [3,5,30]. *Significant difference, $p < 0.01$. doi:10.1371/journal.pone.0088099.g006

and provides significant discrepancies across specific HIV-1 variants and CS, thus improving the available GenBank data for specific-HIV-1 variants consensus sequences.

By using a large dataset of 12,934 sequences from all HIV-1 variants, our study revealed that the CS3, CS5, CS7 and CS8 were the least conserved processing sites across all HIV-1 variants. This finding is in agreement with previous publication using a smaller dataset with 32 subtype C, 34 subtype B and 18 other subtypes sequences [3]. Additional studies are necessary to understand the higher variability in these CS with important roles in the viral cycle. In more detail, CS3 is the first processing site in Gag and GagPol precursors and it is critical for RNA dimer maturation [34]; CS7 is involved in the activation of the viral PR and in the timing and specificity of GagPol cleavage [35]; CS5 is responsible for protein P6^{gag} synthesis which is required for the mature and infectious virion release [36]; CS8 is essential for PR autoprocessing and, it could probably be involved in the correct required PR dimerization [37].

Structural Constrains to CS Variation

Complex interactions of the substrate amino acids within the active site of the viral PR are required for efficient Gag and GagPol cleavage by the PR. HIV-1 PR is only functional in dimeric form and a single monomer is embedded within each precursor. Two individual monomers in different GagPol chains must, therefore, come together to form an embedded dimeric PR, which ultimately cleaves itself into a mature form [37]. HIV-1 maturation requires the recognition by PR of the asymmetric, three-dimensional conformation of the Gag substrate, rather than a particular peptide sequence [38] and, afterwards, PR mediates the cleavage of the HIV-1 structural Gag and GagPol polyproteins by interacting with specific CS [6,39]. Each substrate has a unique structure that differs in amino acid composition [3]. It is thought that these small differences in substrate structure impact affinity for the viral PR and contribute to the highly regulated and ordered stepwise process of maturation in which the individual cleavages occur at different times and rates [3,30]. Additional determinants beyond amino acid sequences and local secondary structures of CS are involved in Gag and GagPol processing [7]. As Gag is conserved, there are constraints on the viability of viral strains with multiple mutations due to the fact that combined mutations are likely to destabilize multiprotein structural interactions that are critical for viral function [40]. Thus, amino acid sequence conservation indicates that the specific amino acids are required to maintain basic structure and function, although other authors have suggested an important role of RNA structure in HIV-1 conservation [33,41]. It is known that physicochemical and structural properties of certain HIV-1 proteins with functional roles in the viral cycle as gp41 can be strongly conserved despite substantial sequence diversity, apparently indicating a delicate balance between evolutionary pressures and the conservation of protein structure [42]. The protein structure, specifically α -helix domains, has been associated with conservation in HIV-1 [33] and is a stable structural element in proteins [43].

Our study reveals which can be the most important CS amino acid sequence for maintaining viral processing by PR and the level of tolerance to amino acid change in each HIV-1 variant. Moreover, the significantly higher conservation observed comparing the late *vs.* the first CS in Gag and GagPol precursors (flanking the PR) would suggest a higher requirement of structural constrains in the last steps of viral processing. Although the aim of our study is purely descriptive, we strongly believe that it can serve as a working tool for research into the better understanding of the CS structure required for a correct cleavage efficacy across

HIV-1 variants and for the design of maturation inhibitors and vaccines targeting CS. Understanding HIV-1 *gag* and *pol* co-evolution [44,45] and the influence of naturally occurring specific-variant polymorphisms at PR [46] in the cleavage process is also crucial for a better interpretation of the biological significance of amino acid changes in CS in the context of a specific HIV-1 variant. Lastly, whether or not the variant specific-residues located in each CS modulate the replicative capacity of the corresponding variant, as was observed for specific natural polymorphisms in the PR in some non-B variants [47], requires further investigation.

It has been suggested that sequences around the CS in Gag are equally conserved as functional motives and sequences targeted by RT inhibitors and are more conserved than non-functional motives [28]. These authors suggested that the amino acid sequences overlapping the CS are immunogenic and, consequently, a vaccine targeting CS could be used for the majority of the world population [28]. Thus, our data on CS conservation across HIV-1 variants could provide useful data to design potential targets for an effective vaccine development against HIV effective for all groups, subtypes and recombinants. Moreover, since mutations within CS have been associated with PI exposure and maturation inhibitor resistance [5,32], our results could potentially provide a better understanding of the role of *gag* in antiretroviral resistance and in the development of future maturation inhibitors [4].

Conclusion

This descriptive study firstly determines the CS conservation degree across most HIV-1 variants and sites in a large dataset composed of 12,934 sequences, inferring the consensus sequences at amino acid level in 11 CS in all Group M subtypes and most CRF and URF, as well as in Groups O, N and P. Our results provide new findings that can help for a better understanding of viral evolution, Gag and GagPol precursors' processing and *gag* structure-function relationships, among others. Our descriptive research could help other researchers in the design of both novel antiretroviral agents acting as maturation inhibitors and for vaccine targeting CS. The biological significance of HIV-1 variant-associated variability found in each processing site in our study needs further future investigation.

Supporting Information

Table S1 HIV-1 variants showing differences with the CS consensus-of-consensuses sequence inferred by GenBank.
(DOC)

Acknowledgments

The authors wish to thank Carolina Fernández McPhee for her assistance in checking the English of the final version of the manuscript and to Israel Pagán for his helpful comments during the manuscript editing.

Author Contributions

Conceived and designed the experiments: AH. Performed the experiments: ET TLD. Analyzed the data: ET TLD. Wrote the paper: ET TLD AH. Created the final figures and tables and submitted the final manuscript: ET. Contributed to figures and tables development: TLD. Reviewed and completed the manuscript and help in the Figure and Tables final design: AH. Wrote the manuscript draft: ET. Contributed to writing the manuscript: TLD. Wrote the final version of manuscript: AH. All authors read and approved the final manuscript.

References

- Swanstrom R, Wills J (1997) Retroviral gene expression. II. Synthesis, processing, and assembly of viral proteins. In: Coffin JM, Hughes SH, Varmus HE, editors. *Retroviruses*. New York: Cold Spring Harbor Laboratory. pp. 263–334.
- Tessmer U, Kräusslich HG (1998) Cleavage of HIV-1 protease from the N-terminally adjacent p6* protein is essential for efficient Gag polyprotein processing and viral infectivity. *J Virol* 72: 3459–3463.
- de Oliveira T, Engelbrecht S, Janse van Rensburg E, Gordon M, Bishop K, et al. (2003) Variability at HIV-1 subtype C protease cleavage sites: an indication of viral fitness? *J Virol* 77: 9422–9430.
- Wahed AA, Freed EO (2012) HIV type 1 Gag as a target for antiviral therapy. *AIDS Res Hum Retroviruses* 28: 54–75.
- Fun A, Wensing AM, Verheyen J, Nijhuis M (2012) Human immunodeficiency virus Gag and protease: partners in resistance. *Retrovirology* 9: 63.
- Adamson CS (2012) Protease-mediated maturation of HIV: inhibitors of protease and the maturation process. *Mol Biol Int* 2012: 1–13.
- Lee SK, Potempa M, Kolli M, Özen A, Schiffer CA, et al. (2012) Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in HIV-1 Gag and Gag-Pro-Pol polyprotein precursors by viral protease. *J Biol Chem* 287: 13279–13290.
- Goodenow MM, Bloom G, Rose SL, Pomeroy SM, O'Brien PO, et al. (2002) Naturally occurring amino acid polymorphisms in HIV-1 Gag p7(NC) and the C-cleavage site impact Gag-Pol processing by HIV-1 protease. *Virology* 292: 137–149.
- Holguín A, Alvarez A, Soriano V (2006) Variability in the P6gag domains of HIV-1 involved in viral budding. *AIDS* 20: 624–627.
- Myint L, Matsuda M, Matsuda Z, Yokomaku Y, Chiba T, et al. (2004) Gag non-cleavage site mutations contribute to full recovery of viral fitness in protease inhibitor-resistant HIV-1. *Antimicrob Agents Chemother* 48: 444–452.
- Doyon L, Payant G, Brakier-Gingras L, Lamarre D (1998) Novel Gag-Pol frameshift site in HIV-1 variants resistant to protease inhibitors. *J Virol* 72: 6146–6150.
- Bally F, Martinez R, Peters S, Sudre P, Telenti A (2000) Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. *AIDS Res Hum Retroviruses* 16: 1209–1213.
- Maguire MF, Guinea R, Griffin P, Macmanus S, Elston RC, et al. (2002) Changes in HIV-1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol* 76: 7398–7406.
- Dam E, Quercia R, Glass B, Descamps D, Launay O, et al. (2009) Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog* 5: e1000345.
- Banké S, Lillemark MR, Gerstoft J, Obel N, Jørgensen LB (2009) Positive selection pressure introduces secondary mutations at Gag cleavage sites in HIV-1 harboring major protease resistance mutations. *J Virol* 83: 8916–8924.
- Nijhuis M, van Maarseveen N, Schipper P (2005) Novel HIV-1 gag based protease drug resistance mechanism caused by an increased processing of the NC/p1 cleavage site. *Antiv Ther* 10: S117.
- Nijhuis M, van Maarseveen NM, Lastere S, Schipper P, Coakley E, et al. (2007) A novel substrate-based HIV-1 protease inhibitor drug resistance mechanism. *PLoS Med* 4: e36.
- Verheyen J, Knops E, Kupfer B, Hamouda O, Somogyi S, et al. (2008) Prevalence of C-terminal gag cleavage site mutations in HIV from therapy-naïve patients. *Journal of Infection* 58: 61–67.
- Ghosh J, Delaunay C, Flandre P, Galimand J, Cohen-Codar I, et al. (2011) Polymorphism in Gag gene cleavage sites of HIV-1 non-B subtype and virological outcome of a first-line lopinavir/ritonavir single drug regimen. *PLoS One* 6: e24798.
- Barrie KA, Perez EE, Lamers SL, Farmerie WG, Dunn BM, et al. (1996) Natural variation in HIV-1 protease, Gag p7 and p6, and protease cleavage sites within gag/pol polyproteins: amino acid substitutions in the absence of protease inhibitors in mothers and children infected by human immunodeficiency virus type 1. *Virology* 219: 407–416.
- Adamson CS, Freed EO (2008) Recent progress in antiretrovirals lessons from resistance. *Drug Discov Today* 13: 424–432.
- Zhang M, Foley B, Schultz AK, Macke JP, Bulla I, et al. (2011) The role of recombination in the emergence of a complex and dynamic HIV epidemic. *Retrovirology* 7: 25.
- Zhuang J, Jetzt AE, Sun G, Yu H, Klarmann G, et al. (2002) Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol* 76: 11273–11282.
- Gao Y, Abreha M, Nelson KN, Baird H, Dudley DM, et al. (2011) Enrichment of intersubtype HIV-1 recombinants in a dual infection system using HIV-1 strain-specific siRNAs. *Retrovirology* 8: 5.
- Peeters M (2000) Recombinant HIV sequences: Their role in the global epidemic. In: Kuiken C, Foley B, Hahn B, Korber B, McCutchan F, Marx P, editors. *Theoretical Biology and Biophysics Group*. Los Alamos NM: National Laboratory. pp. 1–39–1–54.
- Yebra G, de Mulder M, Martín L, Rodríguez C, Labarga P, et al. (2012) Most HIV type 1 non-B infections in the Spanish cohort of antiretroviral treatment-naïve HIV-infected patients (CoRIS) are due to recombinant viruses. *J Clin Microbiol* 50: 407–413.
- Liégeois F, Reteno DG, Mouinga-Ondémé A, Sica J, Rouet F (2013) Short communication: high natural polymorphism in the gag gene cleavage sites of non-B HIV type 1 isolates from Gabon. *AIDS Res Hum Retroviruses* 29: 1179–1182.
- Luo M, Capina R, Daniuk C, Tuff J, Peters H, et al. (2013) Immunogenicity of sequences around HIV-1 protease cleavage sites: Potential targets and population coverage analysis for a HIV vaccine targeting protease cleavage sites. *Vaccine* 31: 3000–3008.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
- Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R (2005) Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-Pol precursor are cleaved by the viral protease at different rates. *Retrovirology* 2: 66.
- Doherty RS, De Oliveira T, Seebregts C, Danaviah S, Gordon M, et al. (2005) BioAfrica's HIV-1 proteomics resource: combining protein data with bioinformatics tools. *Retrovirology* 2: 18.
- Malet I, Roquebert B, Dalban C, Wirdein M, Amellal B, et al. (2007) Association of Gag cleavage sites to protease mutations and to virological response in HIV-1 treated patients. *J Infect* 54: 367–374.
- Snoeck J, Fellay J, Bartha I, Douek DC, Telenti A (2011) Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints. *Retrovirology* 8: 87.
- Shehu-Xhilaga M, Krausslich HG, Pettit S, Swanstrom R, Lee JY, et al. (2001) Proteolytic processing of the p2/nucleocapsid cleavage site is critical for human immunodeficiency virus type 1 RNA dimer maturation. *J Virol* 75: 9156–9164.
- Tomasselli AG, Heinrichson RL (1994) Specificity of retroviral proteases: an analysis of viral and nonviral protein substrates. *Methods Enzymol* 241: 279–301.
- Müller B, Patschinsky T, Kräusslich HG (2002) The late-domain-containing protein p6 is the predominant phosphoprotein of human immunodeficiency virus type 1 particles. *J Virol* 76: 1015–1024.
- Sadiq SK, Noé F, De Fabritius G (2012) Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc Natl Acad Sci U S A* 109: 20449–20454.
- Prabu-Jeyabalan M, Nalivaika E, Schiffer C (2002) Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure* 10: 369–381.
- Kaplan A, Manchester M, Swanstrom R (1994) The activity of the protease of HIV-1 is initiated at the membrane of infected cells before the release of viral proteins and is required for release to occur with maximum efficiency. *J Virol* 68: 6782–6786.
- Dahirel V, Shekhar K, Pereyra F, Miura T, Artyomov M, et al. (2011) Coordinate linkage of HIV evolution reveals regions of immunological vulnerability. *Proc Natl Acad Sci U S A* 108: 11530–11535.
- van der Kuyl AC, Berkhout B (2012) The biased nucleotide composition of the HIV genome: a constant factor in a highly variable virus. *Retrovirology* 9: 92.
- Steckbeck JD, Craigio JK, Barnes CO, Montelaro RC (2011) Highly conserved structural properties of the C-terminal tail of HIV-1 gp41 protein despite substantial sequence variation among diverse clades: implications for functions in viral replication. *J Biol Chem* 286: 27156–27166.
- Richardson J (1981) The anatomy and taxonomy of protein structure. *Adv Protein Chem* 34: 167–339.
- Kozisek M, Henke S, Sasková KG, Jacobs GB, Schuch A, et al. (2012) Mutations in HIV-1 gag and pol compensate for the loss of viral fitness caused by a highly mutated protease. *Antimicrob Agents Chemother* 56: 4320–4330.
- Rossi AH, Rocco CA, Mangano A, Sen L, Aulicino PC (2013) Sequence variability in p6 gag protein and gag/pol coevolution in human immunodeficiency type 1 subtype F genomes. *AIDS Res Hum Retroviruses* 29: 1056–1060.
- Yebra G, de Mulder M, del Romero J, Rodríguez C, Holguín A (2010) HIV-1 non-B subtypes: High transmitted NRTI-resistance in Spain and impaired genotypic resistance interpretation due to variability. *Antiviral Research* 85: 409–417.
- Holguín A, Suñe C, Hany F, Soriano V, Klimkait T (2006) Natural polymorphisms in the protease gene modulate the replicative capacity of non-B HIV-1 variants in the absence of drug pressure. *J Clin Virol* 36: 264–271.

ANEXO V

Otras publicaciones



Enfermedades Infecciosas y Microbiología Clínica

www.elsevier.es/eimc



Original article

Clinical, epidemiological and treatment failure data among HIV-1 non-B-infected patients in the Spanish AIDS Research Network Cohort



Esther Torrecilla García^a, Gonzalo Yebra Sanz^a, Teresa Llacer-Delicado^a, Rafael Rubio García^b, Juan González-García^c, Federico García García^d, José López-Aldegue^e, Víctor Asensi Álvarez^f, África Holguín Fernández^{a,*}, Cohort of the Spanish AIDS Research Network (CoRIS)

^a HIV-1 Molecular Epidemiology Laboratory, Microbiology and Parasitology Department, Hospital Universitario Ramón y Cajal-Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), CIBER en Epidemiología y Salud Pública (CIBERESP), Madrid, Spain

^b HIV Unit, Hospital Universitario 12 de octubre, Madrid, Spain

^c Infectious Diseases Unit, Hospital Universitario La Paz, Madrid, Spain

^d Microbiology Department, Hospital Universitario San Cecilio, Granada, Spain

^e Infectious Diseases Unit, Hospital La Fe, Valencia, Spain

^f Infectious Diseases Unit, Hospital Universitario Central de Asturias, Oviedo, Spain

ARTICLE INFO

Article history:

Received 26 February 2015

Accepted 27 July 2015

Available online 11 September 2015

Keywords:

CoRIS

HIV-1

Non-B variants

Antiretroviral treatment (ART)

Therapeutic failure (TF) to first ART

Epidemiology

ABSTRACT

Introduction: The prevalence of HIV-1 non-B variants is increasing in Spain, showing a higher number of transmitted drug resistance mutations (TDR) since 2002. This study presents the features of non-B-infected patients enrolled in the cohort of antiretroviral treatment (ART) naïve HIV-infected patients included in the Research Network on HIV/AIDS (CoRIS).

Methods: The study includes a selected group of HIV-1 non-B-infected subjects from 670 subjects with pol sequences collected from 2004 to 2008 in the CoRIS cohort. Epidemiological-clinical-virological data were analyzed since cohort entry until October 2011, considering the presence or absence of treatment failure (TF).

Results: Eighty-two non-B-infected subjects with known HIV-1 variants were selected from 2004 to 2008 in the CoRIS cohort, being mainly female, immigrants, infected by recombinant viruses, and by heterosexual route. They had an intermediate TDR rate (9.4%), a high rate of TF (25.6%), of losses to follow-up (35%), of coinfections (32.9%), and baseline CD4+ counts ≥ 350 cells/mm³ (61.8%). Non-B subjects with TF showed higher rates of heterosexual infection (85.7% vs. 69.5%, $p < 0.05$), tuberculosis (30.8% vs. 9.1%, $p = 0.10$) and hepatitis C (23.8% vs. 13.9%, $p = 0.34$) coinfections and lower rates of syphilis (0% vs. 21.9%, $p < 0.05$), and had more frequently received first-line ART including protease inhibitors (PIs) than patients without TF (70% vs. 30%, $p < 0.05$). Interestingly, infection with non-B variants reduced the risk of TDR to nucleoside reverse transcriptase inhibitors and increased the risk to PIs.

Conclusion: HIV-1 non-B-infected patients in Spain had a particular epidemiological and clinical profile that should be considered during their clinical management.

© 2015 Elsevier España, S.L.U. and Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica. All rights reserved.

Datos clínicos, epidemiológicos y fracaso terapéutico en los pacientes VIH-1 infectados por variantes no-B incluidos en la Cohorte de la Red española de Investigación en Sida (CoRIS)

RESUMEN

Palabras clave:

CoRIS

VIH-1

Variantes no-B

Introducción: La prevalencia de variantes no-B del VIH-1 está aumentando en España, mostrando un incremento de las mutaciones de resistencia transmitidas (TDR) desde 2002. Este estudio muestra las características de los pacientes infectados por variantes no-B de la cohorte de infectados por VIH sin tratamiento antirretroviral de la Red de Investigación sobre VIH/SIDA (CoRIS).

* Corresponding author.

E-mail address: africa.holguin@salud.madrid.org (Á. Holguín Fernández).

<http://dx.doi.org/10.1016/j.eimc.2015.07.016>

0213-005X/© 2015 Elsevier España, S.L.U. and Sociedad Española de Enfermedades Infecciosas y Microbiología Clínica. All rights reserved.